

Certificate of Mailing

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: [Mail Stop: Non-Fee Amendment], Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on _____

By: August 27, 2003

Printed: Katherine Sofer

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of: Yue et al.

Title: INTRACELLULAR SIGNALING MOLECULES

Serial No.: 10/018,170 Filing Date: December 11, 2001

Examiner: Steadman, D Group Art Unit: 1652

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF LARS MICHAEL FURNESS

UNDER 37 C.F.R. § 1.132

I, L. MICHAEL FURNESS, a citizen of the United Kingdom, residing at 2 Brookside, Exning, Newmarket, United Kingdom, declare that:

1. I was employed by Incyte Genomics, Inc. (hereinafter "Incyte") as a Director of Pharmacogenomics until December 31, 2001. I am currently under contract to be a Consultant to Incyte Genomics, Inc.

2. In 1984, I received a B.Sc.(Hons) in Biomolecular Science (Biophysics and Biochemistry) from Portsmouth Polytechnic.

From 1985-1987 I was at the School of Pharmacy in London, United Kingdom, during which time I analyzed lipid methyltransferase enzymes using a variety of protein analysis methods, including one-dimensional (1D) and two-dimensional (2D) gel electrophoresis, HPLC, and a variety of enzymatic assay systems.

I then worked in the Protein Structure group at the National Institute for Medical Research until 1989, setting up core facilities for nucleic acid synthesis and sequencing, as well as assisting in programs on protein kinase C inhibitors.

After a year at Perkin Elmer-Applied Biosystems as a technical specialist, I worked at the Imperial Cancer Research Fund between 1990-1992, on a Eureka-funded program collaborating with Amersham Pharmacia in the United Kingdom and CEPH (Centre d'Etude du Polymorphisme Humaine) in Paris, France, to develop novel nucleic acid purification and characterization methods.

In 1992, I moved to Pfizer Central Research in the United Kingdom, where I stayed until 1998, initially setting up core DNA sequencing and then a DNA arraying facility for gene expression analysis in 1993. My work also included bioinformatics and I was responsible for the support of all Pfizer neuroscience programs in the United Kingdom. This then led me into carrying out detailed bioinformatics and wet lab work on the sodium channels, including antibody generation, Western and Northern analyses, PCR, tissue distribution studies, and sequence analyses on novel sequences identified.

In 1998, I moved to Incyte Genomics, Inc., to the Pharmacogenomics group to look at the application of genomics and proteomics to the pharmaceutical industry. In 1999, I was appointed director of the LifeExpress Lead Program which used microarray and protein expression data to identify pharmacologically and toxicologically relevant mechanisms to assist in improved drug design and development.

On December 12, 2001 I founded Nuomics Consulting Ltd., in Exning, U.K., and I am currently employed as Managing Director. Nuomics Consulting Ltd. will be providing expert technical knowledge and advice to businesses around the areas of genomics, proteomics, pharmacogenomics, toxicogenomics and chemogenomics.

3. I have reviewed the specification of a United States patent application that I understand was filed on December 11, 2001 in the names of Yue et al. and was assigned Serial No. 10/018,170 (hereinafter "the Yue '170 application"). Furthermore, I understand that this United States patent application claimed priority to United States Provisional Patent Application Serial No.

60/139,566 filed on June 16, 1999 (hereinafter “the Yue ‘566 application”). The SEQ ID NO:12-encoding polynucleotides were described in the Yue ‘566 application. (Note that the sequences of SEQ ID NO:12 and SEQ ID NO:64 disclosed in the Yue ‘170 application are identical to the sequences referred to as SEQ ID NO:12 and SEQ ID NO:31, respectively, in the Yue ‘566 application). My remarks herein will therefore be directed to the Yue ‘566 patent application, and June 16, 1999, as the relevant date of filing. In broad overview, the Yue ‘566 specification pertains to certain nucleotide and amino acid sequences and their use in a number of applications, including gene and protein expression monitoring applications that are useful in connection with (a) developing drugs (e.g., for the treatment of cancer), and (b) monitoring the activity of drugs for purposes relating to evaluating their efficacy and toxicity.

4. I understand that (a) the Yue ‘170 application contains claims that are directed to a substantially purified polypeptide having the sequence disclosed in the Yue ‘170 application as SEQ ID NO:12 (hereinafter “the SEQ ID NO:12 polypeptide”), and (b) the Patent Examiner has rejected those claims on the grounds that the specification of the Yue ‘170 application does not disclose a substantial, specific and credible utility for the claimed SEQ ID NO:12 polypeptide. I further understand that whether or not a patent specification discloses a substantial, specific and credible utility for its claimed subject matter is properly determined from the perspective of a person skilled in the art to which the specification pertains at the time of the patent application was filed. In addition, I understand that a substantial, specific and credible utility under the patent laws must be a “real-world” utility.

5. I have been asked (a) to consider with a view to reaching a conclusion (or conclusions) as to whether or not I agree with the Patent Examiner’s position that the Yue ‘170 application and its parent, the Yue ‘566 application, does not disclose a substantial, specific and credible “real-world” utility for the claimed SEQ ID NO:12 polypeptide, and (b) to state and explain the bases for any conclusions I reach. I have been informed that, in connection with my considerations, I should determine whether or not a person skilled in the art to which the Yue ‘566 application pertains

on June 16, 1999, would have concluded that the '566 application disclosed, for the benefit of the public, a specific beneficial use of the SEQ ID NO:12 polypeptide in its then available and disclosed form. I have also been informed that, with respect to the "real-world" utility requirement, the Patent and Trademark Office instructs its Patent Examiners in Section 2107 of the Manual of Patent Examining Procedure, under the heading "I. 'Real-World Value' Requirement":

"Many research tools such as gas chromatographs, screening assays, and nucleotide sequencing techniques have a clear, specific and unquestionable utility (e.g., they are useful in analyzing compounds). An assessment that focuses on whether an invention is useful only in a research setting thus does not address whether the specific invention is in fact 'useful' in a patent sense. Instead, Office personnel must distinguish between inventions that have a specifically identified utility and inventions whose specific utility requires further research to identify or reasonably confirm."

6. I have considered the matters set forth in paragraph 5 of this Declaration and have concluded that, contrary to the position I understand the Patent Examiner has taken, the specification of the Yue '566 patent application disclosed to a person skilled in the art at the time of its filing a number of substantial, specific and credible real-world utilities for the claimed SEQ ID NO:12 polypeptide. More specifically, persons skilled in the art on June 16, 1999 would have understood the Yue '566 application to disclose the use of the SEQ ID NO:12 polypeptide as a research tool in a number of gene and protein expression monitoring applications that were well-known at that time to be useful in connection with the development of drugs and the monitoring of the activity of such drugs. I explain the bases for reaching my conclusion in this regard in paragraphs 7-13 below.

7. In reaching the conclusion stated in paragraph 6 of this Declaration, I considered (a) the specification of the Yue '566 application, and (b) a number of published articles and patent documents that evidence gene and protein expression monitoring techniques that were well-known before the June 16, 1999 filing date of the Yue '566 application. The published articles and patent documents I considered are:

- (a) Anderson, N.L., Esquer-Blasco, R., Hofmann, J.-P., Anderson, N.G., A Two-Dimensional Gel Database of Rat Liver Proteins Useful in Gene Regulation and Drug Effects Studies, Electrophoresis, 12, 907-930 (1991) (hereinafter "the Anderson 1991 article") (copy annexed at Tab A);
- (b) Anderson, N.L., Esquer-Blasco, R., Hofmann, J.-P., Mehues, L., Raymackers, J., Steiner, S. Witzmann, F., Anderson, N.G., An Updated Two-Dimensional Gel Database of Rat Liver Proteins Useful in Gene Regulation and Drug Effect Studies, Electrophoresis, 16, 1977-1981 (1995) (hereinafter "the Anderson 1995 article") (copy annexed at Tab B);
- (c) Wilkins, M.R., Sanchez, J.-C., Gooley, A.A., Appel, R.D., Humphery-Smith, I., Hochstrasser, D.F., Williams, K.L., Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It, Biotechnology and Genetic Engineering Reviews, 13, 19-50 (1995) (hereinafter "the Wilkins article") (copy annexed at Tab C);
- (d) Celis, J.E., Rasmussen, H.H., Leffers, H., Madsen, P., Honore, B., Gesser, B., Dejgaard, K., Vandekerckhove, J., Human Cellular Protein Patterns and their Link to Genome DNA Sequence Data: Usefulness of Two-Dimensional Gel Electrophoresis and Microsequencing, FASEB Journal, 5, 2200-2208 (1991) (hereinafter "the Celis article") (copy annexed at Tab D);
- (e) Franzen, B., Linder, S., Okuzawa, K., Kato, H., Auer, G., Nonenzymatic Extraction of Cells from Clinical Tumor Material for Analysis of Gene Expression by Two-Dimensional Polyacrylamide Gel Electrophoresis, Electrophoresis, 14, 1045-1053 (1993) (hereinafter "the Franzen article") (copy annexed at Tab E);
- (f) Bjellqvist, B., Basse, B., Olsen, E., Celis, J.E., Reference Points for Comparisons of Two-Dimensional Maps of Proteins from Different Human Cell Types Defined in a pH Scale Where Isoelectric Points Correlate with Polypeptide Compositions, Electrophoresis, 15, 529-539 (1994) (hereinafter "the Bjellqvist article") (copy annexed at Tab F);
- (g) Large Scale Biology Company Info; LSB and LSP Information; from <http://www.lsbc.com> (2001) (copy annexed at Tab G);

8. Many of the published articles I considered (i.e., at least items (a)-(f) identified in paragraph 7) relate to the development of protein two-dimensional gel electrophoretic techniques for use in protein expression monitoring applications in drug development and toxicology. As I will discuss below, a person skilled in the art who read the Yue '566 application on June 16, 1999 would have understood that application to disclose the SEQ ID NO:12 polypeptide to be useful for a number of protein expression monitoring applications, e.g., in the use of two-dimensional polyacrylamide gel electrophoresis and western blot analysis of tissue samples in drug development and in toxicity testing.

Furthermore, items (a)-(f) establish that protein two-dimensional polyacrylamide gel electrophoresis and western blot analysis were well-known and established methods routinely used in toxicology testing and drug development at the time of filing the Yue '566 application and for several years prior to June 16, 1999. As such, one of ordinary skill in the art would have recognized that the polypeptide of SEQ ID NO:12 could be used in toxicology testing and drug development, irrespective of its biochemical activities.

9. Turning more specifically to the Yue '566 specification, the SEQ ID NO:12 polypeptide is shown at pages 14-15 as one of 38 sequences under the heading "Sequence Listing." The Yue '566 specification specifically teaches that the invention features a "substantially purified" polypeptide, human annexin (INSIG-12) having the amino acid sequence shown in SEQ ID NO:12" (Yue '566 application at p. 3 and Table 2). It further teaches that (a) the identity of the SEQ ID NO:12 polypeptide was determined from a diseased gallbladder tissue cDNA library (GBLANOT01) (Yue '566 application, Table 4), (b) the SEQ ID NO:12 polypeptide is the annexin referred to as "INSIG-12" and is encoded by SEQ ID NO:64, and (c) northern analysis shows that "INSIG-12 is expressed predominantly in cDNA libraries associated with reproductive, gastrointestinal, and nervous system tissues and in tissues associated with cancer and inflammation (Yue '566 application at Table 3).

The Yue '566 application discusses a number of uses of the SEQ ID NO:12 polypeptide in addition to its use in protein expression monitoring applications. I have not fully evaluated these additional uses in connection with the preparation of this Declaration and do not

express any views in this Declaration regarding whether or not the Yue '566 specification discloses these additional uses to be substantial, specific and credible real-world utilities of the SEQ ID NO:12 polypeptide. Consequently, my discussion in this Declaration concerning the Yue '566 application focuses on the portions of the application that relate to the use of the SEQ ID NO:12 polypeptide in gene and protein expression monitoring applications.

10. The Yue '566 application discloses that the polynucleotide sequences disclosed therein, including the polynucleotides encoding the SEQ ID NO:12 polypeptide, are useful as probes in chip based technologies. It further teaches that the chip based technologies can be used "for the detection and/or quantification of nucleic acid or protein" (Yue '566 application at p. 20, lines 17-20).

The Yue '566 application also discloses that the SEQ ID NO:12 polypeptide is useful in other protein expression detection technologies. The Yue '566 application states that "[a] methods for detecting and measuring the expression of INSIG using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS)" (Yue '566 application at p. 20, lines 21-24). Furthermore, the Yue '566 application discloses that "[a] variety of protocols for measuring INSIG, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of INSIG expression. Normal or standard values for INSIG expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to INSIG under conditions suitable for complex formation" (Yue '566 application at p. 31, lines 2-6).

In addition, at the time of filing the Yue '566 application, it was well known in the art that "gene" and protein expression analyses also included two-dimensional polyacrylamide gel electrophoresis (2-D PAGE) technologies, which were developed during the 1980s, and as exemplified by the Anderson 1991 and 1995 articles (Tab A and Tab B). The Anderson 1991 article teaches that a 2-D PAGE map has been used to connect and compare hundreds of 2-D gels of rat liver samples from a variety of studies including regulation of protein expression by various drugs and toxic agents (Tab A at p. 907). The Anderson 1991 article teaches an empirically-determined standard curve fitted

to a series of identified proteins based upon amino acid chain length (Tab A at p. 911) and how that standard curve can be used in protein expression analysis. The Anderson 1991 article teaches that "there is a long-term need for a comprehensive database of liver proteins" (Tab A at p. 912).

The Wilkins article is one of a number of documents that were published prior to the June 16, 1999 filing date of the Yue '566 application that describes the use of the 2-D PAGE technology in a wide range of gene and protein expression monitoring applications, including monitoring and analyzing protein expression patterns in human cancer, human serum plasma proteins, and in rodent liver following exposure to toxins. In view of the Yue '566 application, the Wilkins article, and other related pre-June 1999 publications, persons skilled in the art on June 16, 1999 clearly would have understood the Yue '566 application to disclose the SEQ ID NO:12 polypeptide to be useful in 2-D PAGE analyses for the development of new drugs and monitoring the activities of drugs for such purposes as evaluating their efficacy and toxicity, as explained more fully in paragraph 12 below.

With specific reference to toxicity evaluations, those of skill in the art who were working on drug development in June 1999 (and for many years prior to June 1999) without any doubt appreciated that the toxicity (or lack of toxicity) of any proposed drug they were working on was one of the most important criteria to be considered and evaluated in connection with the development of the drug. They would have understood at that time that good drugs are not only potent, they are specific. This means that they have strong effects on a specific biological target and minimal effects on all other biological targets. Ascertaining that a candidate drug affects its intended target, and identification of undesirable secondary effects (i.e., toxic side effects), had been for many years among the main challenges in developing new drugs. The ability to determine which genes are positively affected by a given drug, coupled with the ability to quickly and at the earliest time possible in the drug development process identify drugs that are likely to be toxic because of their undesirable secondary effects, have enormous value in improving the efficiency of the drug discovery process, and are an important and essential part of the development of any new drug. In fact, the desire to identify and understand toxicological effects using the experimental assays described above led Dr Leigh Anderson to found the Large Scale Biology Corporation in 1985, in order to pursue commercial development of the 2-D electrophoretic protein mapping technology he had developed. In addition, the company focused on

toxicological effects on the proteome as clearly demonstrated by its goals and by its senior management credentials described in company documents (see Tab G at pp. 1, 3, and 5).

Accordingly, the teachings in the Yue '566 application, in particular regarding use of SEQ ID NO:12 in differential gene and protein expression analysis (2-D PAGE maps) and in the development and the monitoring of the activities of drugs, clearly includes toxicity studies and persons skilled in the art who read the Yue '566 application on June 16, 1999 would have understood that to be so.

11. As previously discussed (*supra*, paragraphs 7 and 8), my experience with protein analysis methods in the mid-1980s and the several publications annexed to this Declaration at Tabs A through F evidence information that was available to the public regarding two-dimensional polyacrylamide gel electrophoresis technology and its uses in drug discovery and toxicology testing before the June 16, 1999 filing date of the Yue '566 application. In particular the Celis article stated that "protein databases are expected to foster a variety of biological information.... -- among others, drug development and testing" (See Tab D, p. 2200, second column). The Franzen article shows that 2-D PAGE maps were used to identify proteins in clinical tumor material (See Tab E). The Yue '566 application clearly discloses that expression of INSIG is associated with reproductive, gastrointestinal, and nervous system tissues and tissues associated with cancer and inflammation. (Yue '566 application at Table 3). The Bjellqvist article showed that a protein may be identified accurately by its positional co-ordinates, namely molecular mass and isoelectric point (See Tab F). The Yue '566 application clearly disclosed SEQ ID NO:12 from which it would have been routine for one of skill in the art to predict both the molecular mass and the isoelectric point using algorithms well known in the art at the time of filing.

12. A person skilled in the art on June 16, 1999, who read the Yue '566 application, would understand that application to disclose the SEQ ID NO:12 polypeptide to be highly useful in analysis of differential expression of proteins. For example, the specification of the Yue '566 application would have led a person skilled in the art in June 1999 who was using protein expression

monitoring in connection with working on developing new drugs for the treatment of an cancer, immune disorders, neurological disorders, and gastrointestinal disorders to conclude that a 2-D PAGE map that used the substantially purified SEQ ID NO:12 polypeptide would be a highly useful tool and to request specifically that any 2-D PAGE map that was being used for such purposes utilize the SEQ ID NO:12 polypeptide sequence. Expressed proteins are useful for 2-D PAGE analysis in toxicology expression studies for a variety of reasons, particularly for purposes relating to providing controls for the 2-D PAGE analysis, and for identifying sequence or post-translational variants of the expressed sequences in response to exogenous compounds. Persons skilled in the art would appreciate that a 2-D PAGE map that utilized the SEQ ID NO:12 polypeptide sequence would be a more useful tool than a 2-D PAGE map that did not utilize this protein sequence in connection with conducting protein expression monitoring studies on proposed (or actual) drugs for treating cancer, immune disorders, neurological disorders, and gastrointestinal disorders for such purposes as evaluating their efficacy and toxicity.

I discuss in more detail in items (a)-(b) below a number of reasons why a person skilled in the art, who read the Yue '566 specification in June 1999, would have concluded based on that specification and the state of the art at that time, that SEQ ID NO:12 polypeptide would be a highly useful tool for analysis of a 2-D PAGE map for evaluating the efficacy and toxicity of proposed drugs for cancer, immune disorders, neurological disorders, and gastrointestinal disorders by means of 2-D PAGE maps, as well as for other evaluations:

(a) The Yue '566 specification contains a number of teachings that would lead persons skilled in the art on June 16, 1999 to conclude that a 2-D PAGE map that utilized the substantially purified SEQ ID NO:12 polypeptide would be a more useful tool for protein expression monitoring applications relating to drugs for treating cancer, immune disorders, neurological disorders, and gastrointestinal disorders than a 2-D PAGE map that did not use the SEQ ID NO:12 polypeptide sequence. Among other things, the Yue '566 specification teaches that (i) the identity of the SEQ ID NO:12 polypeptide was determined from a "diseased gallbladder" tissue cDNA library (GBLANOT01) (Yue '566 application, Table 4), (ii) the SEQ ID NO:12 polypeptide is the annexin referred to as INSIG-12, and (iii) INSIG-12 is expressed in various libraries derived from reproductive, gastrointestinal, and nervous system tissues and in tissues associated with cancer and

inflammation (Yue '566 application at Table 3), and therefore "INSIG appears to play a role in cancer, immune disorders, neurological disorders, and gastrointestinal disorders" (Yue '566 application at p. 22; see paragraph 9, *supra*). The substantially purified polypeptide could therefore be used as a control to more accurately gauge the expression of INSIG in the sample and consequently more accurately gauge the affect of a toxicant on expression of the gene.

(b) Persons skilled in the art on June 16, 1999 would have appreciated (i) that the protein expression monitoring results obtained using a 2-D PAGE map that utilized a SEQ ID NO:12 polypeptide would vary, depending on the particular drug being evaluated, and (ii) that such varying results would occur both with respect to the results obtained from the SEQ ID NO:12 polypeptide and from the 2-D PAGE map as a whole (including all its other individual proteins). These kinds of varying results, depending on the identity of the drug being tested, in no way detracts from my conclusion that persons skilled in the art on June 16, 1999, having read the Yue '566 specification, would specifically request that any 2-D PAGE map that was being used for conducting protein expression monitoring studies on drugs for treating cancer, immune disorders, neurological disorders, and gastrointestinal disorders (*e.g.*, a toxicology study or any efficacy study of the type that typically takes place in connection with the development of a drug) utilize the SEQ ID NO:12 polypeptide sequence. Persons skilled in the art on June 16, 1999 would have wanted their 2-D PAGE map to utilize the SEQ ID NO:12 polypeptide sequence because a 2-D PAGE map that utilized protein sequence information the polypeptide (as compared to one that did not) would provide more useful results in the kind of protein expression monitoring studies using 2-D PAGE maps that persons skilled in the art have been doing since well prior to June 16, 1999.

The foregoing is not intended to be an all-inclusive explanation of all my reasons for reaching the conclusions stated in this paragraph 12, and in paragraph 6, *supra*. In my view, however, it provides more than sufficient reasons to justify my conclusions stated in paragraph 6 of this Declaration regarding the Yue '566 application disclosing to persons skilled in the art at the time of its filing substantial, specific and credible real-world utilities for the SEQ ID NO:12 polypeptide.

13. Also pertinent to my considerations underlying this Declaration is the fact that the Yue '566 disclosure regarding the uses of the SEQ ID NO:12 polypeptide for protein expression monitoring applications is not limited to the use of that protein in 2-D PAGE maps. For one thing, the Yue '566 disclosure regarding the technique used in gene and protein expression monitoring applications is broad (Yue '566 application at, e.g., p. 20, lines 16-20 and p. 31, lines 2-9).

In addition, the Yue '566 specification repeatedly teaches that the protein described therein (including the SEQ ID NO:12 polypeptide) may desirably be used in any of a number of long established "standard" techniques, such as ELISA or western blot analysis, for conducting protein expression monitoring studies. See, e.g.:

(a) Yue '566 application at p. 20, lines 21-24 ("Immunological methods for detecting and measuring the expression of INSIG using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS)");

(b) Yue '566 application at p. 31, lines 2-9 ("A variety of protocols for measuring INSIG, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of INSIG expression. Normal or standard values for INSIG expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to INSIG under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of INSIG expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease").

Thus a person skilled in the art on June 16, 1999, who read the Yue '566 specification, would have routinely and readily appreciated that the SEQ ID NO:12 polypeptide disclosed therein would be useful to conduct protein expression monitoring analyses using 2-D PAGE mapping or western blot analysis or any of the other traditional membrane-based protein expression monitoring techniques that were known and in common use many years prior to the filing of the Yue '566 application. For example, a person skilled in the art in June 1999 would have routinely and readily

appreciated that the SEQ ID NO:12 polypeptide would be a useful tool in conducting protein expression analyses, using the 2-D PAGE mapping or western analysis techniques, in furtherance of (a) the development of drugs for the treatment of cancer, immune disorders, neurological disorders, and gastrointestinal disorders, and (b) analyses of the efficacy and toxicity of such drugs.

14. I declare further that all statements made herein of my own knowledge are true and that all statements made herein on information and belief are believed to be true; and further, that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, and that willful false statements may jeopardize the validity of this application and any patent issuing thereon.



L. Michael Furness, B.Sc.

Signed at Exning, United Kingdom
this 18th day of August, 2003

N. Leigh Anderson
Ricardo Esquer-Blasco
Jean-Paul Hofmann
Norman G. Anderson

Large Scale Biology Corporation,
Rockville, MD

genetic form

Contents

1 Introduction.....	907
2 Material and methods	908
2.1 Sample preparation.....	908
2.2 Two-dimensional electrophoresis	909
2.3 Staining	909
2.4 Positional standardization	909
2.5 Computer analysis.....	909
2.6 Graphical data output	910
2.7 Experiment LSBC04.....	910
3 Results and discussion.....	910
3.1 The rat liver protein 2-D map.....	910
3.2 Carbamylated charge standards computed p _i 's and molecular mass standardization	911
3.3 An example of rat liver gene regulation: Cholesterol metabolism	911
3.3.1 MSN 413 (putative cytosolic HMG-CoA synthase) and sets of spots regulated coordinately or inversely	911
3.3.2 MSN 235 and coregulated spots.....	912
3.3.3 An example of an anti-synergistic effect	912
3.3.4 Complexity of the cholesterol synthesis pathway	912
4 Conclusions	912
5 References	912
6 Addendum 1: Figures 1-13	914
7 Addendum 2: Tables 1-4	923
Table 1. Master table of proteins in rat liver database	923
Table 2. Table of some identified proteins	928
Table 3. Computed p _i 's of two sets of carbamylated protein standards: rabbit muscle CPK and human Hb.....	929
Table 4. Computed p _i 's of some known proteins related to measured CPK p _i 's.....	930

A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies

A standard two-dimensional (2-D) protein map of Fischer 344 rat liver (F344MST3) is presented, with a tabular listing of more than 1200 protein species. Sodium dodecyl sulfate (SDS) molecular mass and isoelectric point have been established, based on positions of numerous internal standards. This map has been used to connect and compare hundreds of 2-D gels of rat liver samples from a variety of studies, and forms the nucleus of an expanding database describing rat liver proteins and their regulation by various drugs and toxic agents. An example of such a study, involving regulation of cholesterol synthesis by cholesterol-lowering drugs and a high-cholesterol diet, is presented. Since the map has been obtained with a widely used and highly reproducible 2-D gel system (the Iso-Dalt® system), it can be directly related to an expanding body of work in other lab rat-ries.

1 Introduction

High-resolution two-dimensional electrophoresis of proteins, introduced in 1975 by O'Farrell and others [1-4], has been used over the ensuing 16 years to examine a wide variety of biological systems, the results appearing in more than 5000 published papers. With the advent of computerized systems for analyzing two-dimensional (2-D) gel images and constructing spot databases, it is also possible to plan and assemble integrated bodies of information describing the appearance and regulation of thousands of protein gene products [5, 6]. Creating such databases involves amassing and organizing quantitative data from thousands of 2-D gels, and requires a substantial commitment in technology and resources.

Given the long-term effort required to develop a protein database, the choice of a biological system takes on considerable importance. While *in vitro* systems are ideal for answering many experimental questions, especially in cancer research and genetics, our experience with cell cultures and tissue samples suggests that some *in vivo* approaches could have major advantages. In particular, we have noticed that liver tissue samples from rats and mice appear to show greater quantitative reproducibility (in terms of individual protein expression) than replicate cell cultures. This is perhaps a natural result of the homeostasis maintained in a complete animal vs. the well-known variability of cell cultures, the latter due principally to differences in reagents (e.g., fetal bovine serum), conditions (e.g., pH) and genetic "evolution" of cell lines while in culture. It is also more difficult to generate adequate amounts of protein from cell culture systems (particularly with attached cells), forcing the investigator to resort to radioisotope-based or silver-based stain-detection methods. While these methods are more sensitive (sometimes much more sensitive) than the Coomassie Brilliant Blue (CBB) stain typically used for protein detection in "large" protein samples, they are generally more variable, more labor-intensive and, in the case of radiographic methods, may generate highly "noisy" images, due to the properties of the films used. By contrast, large protein samples can easily be prepared from liver using urea/Nonidet P-40 (NP-40) s. lubilization and stained with CBB, which has the advantage of being easily reproducible [8]. Finally, there remains the question of the "truthfulness" of many *in vitro* systems as compared to their *in vivo* analogs; how great are the changes caused by the introduction into a cul-

Correspondence: Dr. N. Leigh Anderson, Large Scale Biology Corporation, 9620 Medical Center Drive, Rockville, MD 20850, USA

Abbreviations: CBB, Coomassie Brilliant Blue; CPK, creatine phosphokinase; 2-D, two-dimensional; IEF, isoelectric focusing; MSN, master spot number; NP-40, Nonidet P-40. SDS, sodium dodecyl sulfate

CVCH Verlagsgesellschaft mbH, D-6940 Weinheim, 1991

0173-0835/91/1111-0907 \$3.50+25/0

ture and the associated shift to strong selection for growth, and how do these affect experimental outcomes? Hence the apparent advantages of *in vitro* systems, in terms of experimental manipulation, may be counterbalanced by other factors relating to 2-D data quality.

There is a second important class of reasons for exploring the use of an *in vivo* biological system such as the liver. Historically, there have been two broad approaches to the mechanistic dissection of biochemical processes in intact cellular systems: genetics (a search for informative mutants) and the use of chemical agents (drugs and chemical toxins). Both approaches help us to understand complex systems by disrupting some specific functional element and showing us the result. With the development of techniques for genetic manipulation and cloning, the genetic approach can be effectively applied either *in vitro* or *in vivo*, although the *in vitro* route is usually quicker. The chemical approach can also be applied to either sort of biological system; here, however, the bulk of consistently acquired information is in experimental animals (rats and mice). While most biologists know a short list of compounds having specific, experimentally useful effects (e.g., inhibitors of protein synthesis, ionophores, polymerase inhibitors, channel blockers, nucleotide analogs, and compounds affecting polymerization of cytoskeletal proteins), there is a much larger number of interesting chemically-induced effects, most of them characterized by toxicologists and pharmacologists in rodent systems. Just as a thorough genetic analysis would involve saturating a genome with mutations, it is possible to imagine a saturating number of drugs, the analysis of whose actions would reveal the complete biochemistry of the cell. While organized drug discovery efforts usually target specific desired effects, the nature of the process, with its dependence on screening large numbers of compounds, necessarily produces many unanticipated effects. It is therefore reasonable to suppose that the required broad range of compounds necessary to achieve "biochemical saturation" may be forthcoming; in fact, it may already exist among the hundreds of thousands of compounds that failed to qualify as drugs.

Among organs, the liver is an obvious choice for the study of chemical effects because of its well-known plasticity and responsiveness. The brain appears to be quite plastic (e.g. [7]), but it is a complicated mixture of cell types requiring skillful dissection for most experiments. The kidney, while quite responsive, also presents a potentially confounding mixture of cell types. The liver, by contrast, is made up of one predominant cell type which is easy to solubilize: the hepatocyte, representing more than 95% of its mass. Most importantly, the liver performs many homeostatic functions that require rapid modulation of gene expression. It appears that most chemical agents tested affect gene expression in the liver at some dosage (N. Leigh Anderson, unpublished observations), an interesting contrast to our earlier work with lymphocytes, for example, which seem to be much less responsive. Such results conform to the expectation that cells with a homeostatic, physiological role should be more plastic than cells differentiated for a purpose dependent on the action of a limited number of specific genes.

The liver also allows the parallels between *in vitro* and *in vivo* systems to be examined in detail. Significant progress

has been made in the development of mouse, rat and human hepatocyte culture systems, as well as in precision-cut tissue slices. Using such an array of techniques, it is possible to assemble a matrix of mammalian systems including mouse and rat *in vivo* on one level and mouse, rat and man *in vitro* on a second level, and to compare effects between species and between systems. This approach allows us to draw informed conclusions regarding the biochemical "universality" of biological responses among the mammals, and to offer some insight into the validity of *in vitro* approaches for toxicological screening. We believe this will be necessary if *in vitro* alternatives are to achieve widespread usage in government-mandated safety testing of drugs, consumer products and industrial and agricultural chemicals.

A number of interesting studies have been published using 2-D mapping to examine effects in the rodent liver. A number of investigators have made use of the technique to screen for existing genetic variants [8–11] or induced mutations [12–14], mainly in the mouse. This work builds on the wealth of genetic information available on the mouse and its established position as a mammalian mutation-detection system. While some studies of chemical effects have been undertaken in the mouse [15–17], most have used the rat [18–23]. The examination of the cytochrome p-450 system, in particular, has been carried out almost exclusively on the rat [24, 25].

These considerations lead us to conclude that rodent liver offers the best opportunity to systematically examine an array of gene regulation systems, and ultimately to build a predictive model of large-scale mammalian gene control. The basic underlying foundation of such a project is a reliable, reproducible master 2-D pattern of liver, to which ongoing experimental results can be referred. In this paper, we report such a master pattern for the acidic and neutral proteins of rat liver (pattern F344MST3). In future, this master will be supplemented by maps of basic proteins, and analogous maps of mouse and human liver.

2 Materials and methods

2.1 Sample preparation

Liver is an ideal sample material for most biochemical studies, including 2-D analysis. A sample is taken of approximately 0.5 g of tissue from the apical end of the left lobe of the liver. Solubilization is effected as rapidly as practical; a delay of 5–15 min appears to cause no major alteration in liver protein composition if the liver pieces are kept cold (e.g., on ice) in the interim. In the solubilization process, the liver sample is weighed, placed in a glass homogenizer (e.g., 15 mL Wheaton); 8 volumes of solubilizing solution*

* The solubilizing solution is composed of 2% NP-40 (Sigma), 9 M urea (analytical grade, e.g., BDH or Bio-Rad), 0.5% dithiothreitol (DTT; Sigma) and 2% carrier ampholytes (pH 9–11 LKB; these come as a 20% stock solution, so 2% final concentration is achieved by making the final solution 10% 9–11 Amphotoline by volume). A large batch of solubilizer (several hundred mL) is made and stored frozen at –80°C in aliquots sufficient to provide enough for one day's estimated sample preparation requirement. The solution is never allowed to become warmer than room temperature at any stage during preparation or thawing for use, since heating of concentrated urea solutions can produce contaminants that covalently modify proteins producing artifactual charge shifts. Once thawed, any unused solubilizer is discarded.

added (i.e., 4 mL per 0.5 g tissue) and the mixture is homogenized using first the loose- and then the tight-fitting glass pestle. This takes approximately 5 strokes with the pestle and is carried out at room temperature because it would crystallize out in the cold. Once the liver sample is thoroughly homogenized in the solubilizer, it is assumed that all the proteins are denatured (by the chaotropic effect of the urea and NP-40 detergent) and the enzymes inactivated by the high pH (~9.5). Therefore these samples may be kept at room temperature until they can be centrifuged frozen as a group (within several hours of preparation). The samples are centrifuged for 6×10^5 g min (e.g., 500 000 g for 12 min using a Beckman TL-100 centrifuge). The centrifuge rotor is maintained at just below room temperature (e.g., 15–20°C), but not too cold, so as to prevent the precipitation of urea. The centrifuge of choice is a Beckman L-100 because of the sample tube sizes available, but any ultracentrifuge accepting smallish tubes will suffice. When an appropriate centrifuge is not available near the site of sample preparation, samples can be frozen at -80°C and thawed prior to centrifugation and collection of supernatants. Each supernatant is carefully removed following centrifugation and aliquoted into at least 4 clean tubes for storage. This is done by transferring all the supernatant to one clean tube, mixing this gently (to assure homogeneous composition) and then dividing it into 4 aliquots. The aliquots are frozen immediately at -80°C. These multiple aliquots can provide insurance against a failed run or a freezer breakdown.

2. Two-dimensional electrophoresis

Sample proteins are resolved by 2-D electrophoresis using the 20×25 cm Iso-Dalt³ 2-D gel system ([26–29]; produced by LSB and by Hoefer Scientific Instruments, San Francisco) operating with 20 gels per batch. All first-dimensional isoelectric focusing (IEF) gels are prepared using the same single standardized batch of carrier ampholytes BDH 4–8A in the present case, selected by LSB's batch-testing program for rat and mouse database work**). A 10 μ L sample of solubilized liver protein is applied to each gel, and the gels are run for 33 000 to 34 500 volt-hours using a progressively increasing voltage protocol implemented by a programmable high-voltage power supply. An "Angeline" computer-controlled gradient-casting system (produced by LSB) is used to prepare second-dimensional sodium dodecyl sulfate (SDS) polyacrylamide gradient slab gels in which the top 5% of the gel is 11%T acrylamide, and the lower 95% of the gel varies linearly from 11% to 18%T.

This system has recently been modified so as to employ a commercially available 30.8%T acrylamide/*N,N'*-methylenebisacrylamide prepared solution (thus avoiding the handling of the solid acrylamide monomer) and three additional stock solutions: buffer (made from Sigma pre-set Tris), persulfate and *N,N,N',N'*-tetramethylethylenediamine (TEMED). Each gel is identified by a computer-printed filter paper label polymerized into the lower left corner of the gel. First-dimensional IEF tube gels are loaded

*This material (succeeding certified batches of which are available from Hoefer Scientific Instruments) has the most linear pH gradient produced by any ampholyte tested except for the Pharmacia wide range which has an unacceptable tendency to bind high-molecular weight acidic proteins, causing them to streak.

directly (as extruded) onto the slab gels without equilibration, and held in place by polyester fabric wedges ("Wedges", produced by LSB) to avoid the use of hot agarose. Second-dimensional slab gels are run overnight, in groups of 20, in cooled DALT tanks (10°C) with buffer circulation. All run parameters, reagent source and lot information, and notations of deviation from expected results are entered by the technician responsible on a detailed, multi-page record of the experiment.

2.3 Staining

Following SDS-electrophoresis, slab gels are stained for protein using a colloidal Coomassie Blue G-250 procedure in covered plastic boxes, with 10 gels (totalling approximately 1 L of gel) per box. This procedure (based on the work of Neuhoff [30, 31]) involves fixation in 1.5 L of 50% ethanol and 2% phosphoric acid for 2 h, three 30 min washes, each in 2 L of cold tap water, and transfer to 1.5 L of 34% methanol, 17% ammonium sulfate and 2% phosphoric acid for 1 h, followed by the addition of a gram of powdered Coomassie Blue G-250 stain. Staining requires approximately 4 days to reach equilibrium intensity, whereupon gels are transferred to cool tap water and their surfaces rinsed to remove any particulate stain prior to scanning. Gels may be kept for several months in water with added sodium azide. The water washes remove ethanol that would dissolve the stain (and render the system noncolloidal, with high backgrounds). The concentrated ammonium sulfate and methanol solution is diluted by equilibration with the water volume of the gels to automatically achieve the correct final concentrations for colloidal staining. Practical advantages of this staining approach can be summarized as follows: (i) the low, flat background makes computer evaluation of small spots (max OD < 0.02) possible, especially when using laser densitometry; (ii) up to 1500 spots can be reliably detected on many gels (e.g., rat liver) at loadings low enough to preserve excellent resolution; and (iii) reproducibility appears to be very good: at least several hundred spots have coefficients of reproducibility less than 15%. This value is at least as good as previous CBB methods, and significantly better than many silver stain systems.

2.4 Positional standardization

The carbamylated rabbit muscle creatine phosphokinase (CPK) standards [32] are purchased from Pharmacia and BDH. Amino acid compositions, and numbers of residues present in proteins used for internal standardization, are taken from the Protein Identification Resource (PIR) sequence database [33].

2.5 Computer analysis

Stained slab gels are digitized in red light at 134 micron resolution, using either a Molecular Dynamics laser scanner (with pixel sampling) or an Eikonix 78/99 CCD scanner. Raw digitized gel images are archived on high-density DAT tape (or equivalent storage media) and a greyscale videotape prepared from the raw digital image as hard-copy backup of the gel image. Gels are processed using the Kepler⁴ software system (produced by LSB), a commercially available workstation-based software package built on

some of the principles of the earlier TYCHO system [34-41]. Procedure PROC008 is used to yield a spotlist giving position, shape and density information for each detected spot. This procedure makes use of digital filtering, mathematical morphology techniques and digital masking to remove the background, and uses full 2-D least-squares optimization to refine the parameters of a 2-D Gaussian shape for each spot. Processing parameters and file locations are stored in a relational database, while various log files detailing operation of the automatic analysis software are archived with the reduced data. The computed resolution and level of Gaussian convergence of each gel are inspected and archived for quality control purposes.

Experiment packages are constructed using the Kepler experiment definition database to assemble groups of 2-D patterns corresponding to the experimental groups (e.g., treated and control animals). Each 2-D pattern is matched to the appropriate "master" 2-D pattern (pattern F344MST3 in the case of Fischer 344 rat liver), thereby providing linkage to the existing rodent protein 2-D databases. The software allows experiments containing hundreds of gels to be constructed and analyzed as a unit, with up to 100 gels displayed on the screen at one time for comparative purposes and multiple pages to accommodate experiments of > 1000 gels. For each treatment, proteins showing significant quantitative differences vs. appropriate controls are selected using group-wise statistical parameters (e.g., Student's *t*-test, Kepler[®] procedure STUDENT). Proteins satisfying various quantitative criteria (such as $P < 0.001$ difference from appropriate controls) are represented as highlighted spots onscreen or on computer-plotted protein maps and stored as spot populations (i.e., logical vectors) in a liver protein database. Quantitative data (spot parameters, statistical or other computed values) are stored as real-valued vectors in the database. Analysis of coregulation is performed using a Pierson product-moment correlation (Kepler procedure CORREL) to determine whether groups of proteins are coordinately regulated by any of the treatments. Such groups can be presented graphically on a protein map, and reported together with the statistical criteria used to assess the level of coregulation. Multivariate statistical analysis (e.g., principal components' analysis) is performed on data exported to SAS (SAS Institute).

2.6 Graphical data output

Graphical results are prepared in GKS and translated within Kepler[®] into output for any of a variety of devices. Linedrawing output is typically prepared as Postscript and printed on an Apple Laserwriter. Detailed maps presented here have been generated using an ultra-high-resolution Postscript-compatible Linotronic output device. Greyscale graphics are reproduced from the workstation screen using a Seikosha videotriplate. Patterns are shown in the standard orientation, with high molecular mass at the top and acidic proteins to the left.

2.7 Experiment LSBC04

In the study described here 12-week-old Charles River male F344 rats were used. Diets were prepared at LSB, based on a Purina 5755M Basal Purified Diet. Lovastatin and cholestyramine were obtained as prescription pharma-

ceuticals, ground and mixed with the diet at concentrations of 0.075% and 1%, respectively. The high cholesterol diet was Purina 5801M-A (5% cholesterol plus 1% sodium cholate in the control diet). Animal work was carried out by Microbiological Associates (Bethesda, MD). Animals were acclimatized for one week on the control diet, fed test or control diets for one week, and sacrificed on day 8. Average daily doses of lovastatin and cholestyramine in appropriate groups were 37 mg/kg/day and 5 g/kg/day, respectively, based on the weight of the food consumed. Liver samples were collected and prepared for 2-D electrophoresis according to the standard liver protocol (homogenization in 8 volumes of 9 M urea, 2% NP-40, 0.5% dithiothreitol, 2% LKB pH 9-11 carrier ampholytes, followed by centrifugation for 30 min at 80 000 $\times g$). Kidney, brain and plasma samples were frozen. Gels were run as described above, and the data was analyzed using the Kepler[®] system. Gels were scaled, to remove the effect of differences in protein loading, by setting the summed abundances of a large number of matched spots equal for each gel (linear scaling).

3 Results and discussion

3.1 The rat liver protein 2-D map

F344MST3 is a standard 2-D pattern of rat liver proteins, based on the Fischer 344 strain. This pattern was initiated from a single 2-D gel and extensively edited in an experiment comparing it to a range of protein loads, so as to include both small spots and well-resolved representations of high-abundance spots. More than 700 rat liver 2-D patterns have been matched to F344MST3 in a series of drug effects and protein characterization experiments, and numerous new spots (induced by specific drugs, for instance) have been added as a result. A modified version including additional spots present in the Sprague-Dawley outbred rat has also been developed (data not shown). Figure 1 shows a greyscale representation and Fig. 2 a schematic plot of the master pattern. More than 1200 spots are included, most of which are visible on typical gels loaded with 10 μ L of solubilized liver protein prepared by the standard method and stained with colloidal Coomassie Blue. Master spot numbers (MSN's) have been assigned to all proteins, and appear in the following figures, each showing one quadrant of the pattern. Figure 3 shows the upper left (acidic, high molecular mass) quadrant, Fig. 4 the upper right (basic, high molecular mass) quadrant, Fig. 5 the lower left (acidic, low molecular mass) quadrant, and Fig. 6 the lower right (basic, low molecular mass) quadrant. The quadrants overlap as an aid to moving between them. The gel position (in 100 micron units), isoelectric point (relative to the CPK internal p/standards) and SDS molecular mass (from the calibration curve in Fig. 8) are listed for each spot (Table 1). Because of the precision of the CPK-p/values, these parameters can be used to relate spot locations between gel systems more reliably than using p/measurements expressed as pH. A major objective of current studies is the identification of all major spots corresponding to known liver proteins, as well as rigorous definitions of subcellular organelle contents. Of particular interest to us is the parallel development of identifications in the rat and mouse liver maps, allowing detailed comparisons of gene expression effects in the two systems. The results of these studies will be presented systematically in a later edition of this database.

we include here a useful series of 22 orienting identifications as an aid to other users of the rat liver pattern (Table 1).

2. Carbamylated charge standards, computed p*is* and molecular mass standardization

We have previously shown that the use of a system of close-spaced internal p*is* markers (made by carbamylating a basic protein) offers an accurate and workable solution to the problem of assigning positions in the p*is*/dimension [32]. The same system, based on 36 protein species made by carbamylating rabbit muscle CPK, has been used here to assign p*is* to most rat liver acidic and neutral proteins. The standards were coelectrophoresed with total liver proteins, and the standard spots added to a special version of the hamster pattern F344MST3. The gel X-coordinates of all liver protein spots lying within the CPK charge train were then transformed into CPK p*is* positions by interpolation between the positions of immediately adjacent standards (Table 1) using a Kepler² vector procedure.

It has proven possible to compute fairly accurate p*is* values for many proteins from the amino acid composition [42]. We have attempted here to test a further elaboration of this approach, in which we computed p*is*s for the CPK standards themselves, based on our knowledge of the rabbit muscle CPK sequence and the fact that adjacent members of the charge train typically differ by blockage of one additional lysine residue (Table 3). We compared these values to similar computed p*is*s for an additional set of carbamylated standards made from human hemoglobin beta chains and a series of rat liver and human plasma proteins of known position and sequence (Fig. 7, Table 4). The result demonstrates good concordance between these systems. Two proteins show significant deviations: liver fatty-acid binding protein (FABP; #1 in Table 4) and protein disulphide isomerase (#20 in the table). The FABP spot present on F344MST3 may represent a charge-modified version of a more basic parent spot closer to the expected p*is*, not resolved in the IEF/SDS gel. Of particular importance is the fact that, by comparing computed p*is*s of sequenced but unlocated proteins with the CPK p*is*s, we can assign a probable gel location without making any assumptions regarding the actual gel pH gradient. This offers a useful shortcut, given the vagaries of pH measurement on small diameter IEF gels. We have used this approach to compute the CPK p*is*s of all rat and mouse proteins in the PIR sequence database, as an aid to protein identification (data not shown).

In order to standardize SDS molecular weight (SDS-MW), we have used a standard curve fitted to a series of identified proteins (Fig. 8). Rather than using molecular mass *per se*, we have elected to use the number of amino acids in the polypeptide chain, as perhaps a better indication of the length of the SDS-coated rod that is sieved by the second dimension slab. The resulting values were multiplied by 12 (the weighted average mass of amino acids in sequenced proteins) to give predicted molecular masses. Because we use gradient slabs, we have not constrained the fit of the curve to conform to any predetermined model; rather we tried many equations and selected the best using the program "Tablecurve" on a PC. The equation chosen was $y = ax^2 + bx + c/x^2$, where y is the number of residues, x is the gel

Y-coordinate, a is 511.83, b is -0.2731 and c is 33183801. The resulting fit appears to be fairly good over a broad range of molecular mass.

3.3 An example of rat liver gene regulation: Cholesterol metabolism

Experiment LSBC04 was designed as a small-scale test of the regulation of cholesterol metabolism *in vivo* by three agents included in the diet: lovastatin (Mevacor², an inhibitor of HMG-CoA reductase); cholestyramine (a bile acid sequestant that has the effect of removing cholesterol from the gut-liver recirculation); and cholesterol itself. The first two agents should lower available cholesterol and the third should raise it, allowing manipulation of relevant gene expression control systems in both directions. Such an experiment offers an interesting test of the 2-D mapping system since most of the pathway enzymes are present in low abundance, many are membrane-bound and difficult to solubilize, and the pathway itself is complex. Approximately 1000 proteins were separated and detected in liver homogenates. Twenty-one proteins were found to be affected by at least one treatment, and these could be divided into several coregulated groups.

3.3.1 MSN 413 (putative cytosolic HMG-CoA synthase) and sets of spots regulated coordinately or inversely

One group of spots (including a spot assigned to the cytosolic HMG-CoA synthase, MSN 413) showed the expected increase in abundance with lovastatin or cholestyramine, the synergistic further increase with lovastatin and cholestyramine, and a dramatic decrease with the high cholesterol diet. Spot number 413 is the most strongly regulated protein in the present experiment, showing a 5- to 10-fold induction after a 1 week treatment with 0.075% lovastatin and 1% cholestyramine in the diet (Figs. 9 and 10). Its expression follows precisely the expectation for an enzyme whose abundance is controlled by the cholesterol level; it is progressively increased from the control levels by cholestyramine, lovastatin and lovastatin plus cholestyramine, and it sinks below the threshold of detection in animals fed the high cholesterol diet. This spot has been tentatively identified as the cytosolic HMG-CoA synthase, based on a reaction with an antiserum to that protein provided by Dr. Michael Greenspan at Merck Sharp & Dohme Research Laboratories. This enzyme lies immediately before HMG-CoA reductase in the liver cholesterol biosynthesis pathway, and is known to be co-regulated with it. Spot 413 has an SDS molecular weight of about 54 000 and a CPK p*is* of -11.4, in reasonably close agreement with a molecular weight of 57300 and a CPK p*is* of -15.7 computed from the known sequence of the hamster enzyme [43].

Using a classical product-moment correlation test (Kepler procedure CORREL), a series of five additional spots was found to be coregulated with 413. The level of correlation was exceedingly high (> 95%). Two of these, 1250 and 933, are at similar molecular weights and approximately one charge more acidic than 413 (Fig. 9), indicating that they may be covalently modified forms of the 413 polypeptide. This suspicion is strengthened by the observation that both spots are also stained by the antibody to cytosolic HMG-CoA synthase. The remaining three correlated spots appear

to comprise an additional related pair (1253 and 1001) of around 40 kDa and a single spot (1119) of around 28 kDa. Because these two presumed proteins are present at substantially lower abundances than 413, and because the cytosolic HMG-CoA synthase is reported to consist of only one type of polypeptide, they are likely to represent other, very tightly coregulated enzymes. A second group of six spots was selected based on a regulatory pattern close to the inverse of that for spot 413 (MSN's 34, 79, 178, 182, 204, 347; data not shown). For these proteins, the lowest level of expression occurs with exposure to lovastatin plus cholestyramine and the highest level upon exposure to the high-cholesterol diet. Spots 182 and 79 are highly correlated and lie about one charge apart at the same molecular weight; they may thus be isoforms of a single protein. The other four spots probably represent additional enzymes or subunits.

3.3.2 MSN 235 and coregulated spots

A third group of five spots, mainly comprised of mitochondrial proteins including putative mitochondrial HMG-CoA synthase spots, showed a modest induction by lovastatin alone, but little or no effect with any of the other treatments (including the combination of lovastatin and cholestyramine; Fig. 12). This result is intriguing because lovastatin was expected to affect only the regulation of enzymes of cholesterol synthesis, which is entirely extra-mitochondrial. Three of the spots (235, 134, 144) form a closely-packed triad at approximately 30 kDa, and are likely to represent isoforms of one protein. All three spots are stained by an antibody to the mitochondrial form of HMG-CoA synthase obtained from Dr. Greenspan. Subcellular fractionation indicates a mitochondrial location. The other two spots (633 at about 38 kDa and 724 at about 69 kDa) are each present at lower abundance than the members of the triad.

3.3.3 An example of an anti-synergistic effect

A sixth spot (367) shows strong induction by lovastatin (two- to threefold), and about half as much induction with lovastatin plus cholestyramine, but without sharing the animal-animal heterogeneity pattern of the 235-set (Fig. 13). This protein is also mitochondrial, and represents the clearest example of an anti-synergistic effect of lovastatin and cholestyramine. The existence of such an effect demonstrates that lovastatin and cholestyramine do not act exclusively through the same regulatory pathway.

3.3.4 Complexity of the cholesterol synthesis pathway

Taken together, these results suggest that treatment with lovastatin alone can affect both cytosolic and mitochondrial pathways using HMG-CoA, while cholestyramine, on the other hand, either alone or in combination with lovastatin, produces a strong effect on the putative cytosolic pathway, but little or no effect on the putative mitochondrial pathway. An explanation for this difference may lie in lovastatin's effect on levels of HMG-CoA and related precursor compounds that are exchanged between the cytosol and the mitochondrion, whereas cholestyramine should affect only the cytosolic pathways directly controlled by cholesterol and bile acid levels. It remains to be explained why some

proteins of the putative mitochondrial pathway are so much more variable in their expression in all groups. An examination of all the coregulated groups suggests that quantitative statistical techniques can extract a wealth of interesting information from large sets of reproducible gels. The abundance of spots in the 413 coregulation group, for example, shows an amazing level of concordance in their relative expression among the five individuals of the lovastatin and cholestyramine treatment group. This effect is not due to differences in total protein loading, since they have already been removed by scaling, and since proteins with quite different regulation patterns can be demonstrated (e.g., Fig. 13). Such effects raise the possibility that many gene coregulation sets may be revealed through the study of a sufficiently large population of control animals (*i.e.*, without any experimental manipulation). This approach, exploiting natural biological variation in protein expression instead of drug effects, offers an important incentive for the construction of a large library of control animal patterns.

4 Conclusions

Because of the widespread use of rat liver in both basic biochemistry and in toxicology, there is a long-term need for a comprehensive database of liver proteins. The rat liver master pattern presented here has proven to be an accurate representation of this system, having been matched to more than 700 gels to date. As the number of proteins identified and the number of compounds tested for gene expression effects grows, we expect this database to contribute valuable insights into gene regulation. Its practical utility in several areas of mechanistic toxicology is already being demonstrated.

Received September 11, 1991

5 References

- [1] O'Farrell, P. J. *J. Biol. Chem.* 1975, 250, 4007-4021.
- [2] Klose, J. *Humangenetik* 1975, 26, 231-243.
- [3] Scheele, G. A. *J. Biol. Chem.* 1975, 250, 5375-5385.
- [4] Iborra, G. and Buhler, J. M. *Anal. Biochem.* 1976, 74, 503-511.
- [5] Anderson, N. G. and Anderson, N. L. *Behring Inst. Mitt.* 1979, 63, 169-210.
- [6] Anderson, N. G. and Anderson, N. L. *Clin. Chem.* 1982, 28, 739-745.
- [7] Heydorn, W. E., Creed, G. J. and Jacobowitz, D. M. *J. Pharmacol. Exp. Therap.* 1984, 229, 622-628.
- [8] Anderson, N. L., Nance, S. L., Tollaksen, S. L., Giere, F. A. and Anderson, N. G. *Electrophoresis* 1985, 6, 592-599.
- [9] Racine, R. R. and Langley, C. H. *Biochem. Genet.* 1980, 18, 185-197.
- [10] Klose, J. *Mol. Evol.* 1982, 18, 315-328.
- [11] Neel, J. V., Baier, L., Hanash, S. and Erickson, R. P. *J. Hered.* 1985, 76, 314-320.
- [12] Marshall, R. R., Raj, A. S., Grant, F. J. and Heddle, J. A. *Can. J. Genet. Cytol.* 1983, 25, 457-466.
- [13] Taylor, J., Anderson, N. L., Anderson, N. G., Gemmell, A., Giometti, C. S., Nance, S. L. and Tollaksen, S. L., in: Dunn, M. J. (Ed.), *Electrophoresis '86*, Verlag Chemie, Weinheim 1986, pp. 583-587.
- [14] Giometti, C. S., Gemmell, M. A., Nance, S. L., Tollaksen, S. L. and Taylor, J., *J. Biol. Chem.* 1987, 262, 12764-12767.
- [15] Anderson, N. L., Giere, F. A., Nance, S. L., Gemmell, M. A., Tollaksen, S. L. and Anderson, N. G., in: Gaiteau, M.-M. and Siest, G. (Eds.), *Progrès Récents en Electrophorèse Bidimensionnelle*, Presses Universitaires de Nancy, Nancy 1986, pp. 253-260.
- [16] Anderson, N. L., Swanson, M., Giere, F. A., Tollaksen, S., Gemmell, A., Nance, S. L. and Anderson, N. G., *Electrophoresis* 1986, 7, 44-48.

- Anderson, N. L., Giere, F. A., Nance, S. L., Gemmell, M. A., Tollakson, S. L. and Anderson, N. G., *Fundam. Appl. Toxicol.* 1987, 8, 39-50.
- Anderson, N. L., in: *New Horizons in Toxicology*, Eli Lilly Symposium, 1991, in press.
- Antoine, B., Rahimi-Pour, A., Sieck, G., Magdalou, J. and Galteau, M. M., *Cell. Biochem. Funct.* 1987, 5, 217-231.
- Elliott, B. M., Ramasamy, R., Stoward, M. D. and Spragg, S. P., *Biochim. Biophys. Acta* 1986, 876, 135-140.
- Huber, B. E., Heilman, C. A., Wirth, P. J., Miller, M. J. and Thorgeirsson, S. S., *Hepatology* 1986, 4, 206-219.
- Wirth, P. J. and Vesterberg, O., *Electrophoresis* 1988, 9, 47-53.
- Witzmann, F. A. and Parker, D. N., *Toxicol. Lett.* 1991, 57, 29-36.
- Rampersaud, A., Waxman, D. J., Ryan, D. E., Levin, W. and Walz, F. G., Jr., *Arch. Biochem. Biophys.* 1985, 243, 174-183.
- Vlasuk, G. P. and Walz, F. G., Jr., *Anal. Biochem.* 1980, 105, 112-120.
- Anderson, N. G. and Anderson, N. L., *Anal. Biochem.* 1978, 85, 331-340.
- Anderson, N. L. and Anderson, N. G., *Anal. Biochem.* 1978, 85, 341-354.
- Anderson, L., Hofmann, J.-P., Anderson, E., Walker, B. and Anderson, N. G., in: Endler, A. T. and Hanzlik, S. (Eds.), *Two-Dimensional Electrophoresis*, VCH Verlagsgesellschaft, Weinheim 1989, pp. 288-297.
- Anderson, L., *Two-Dimensional Electrophoresis: Operation of the ISO-DALT® System*, Large Scale Biology Press, Washington, DC 1988, ISBN 0-945532-00-8, 170pp.
- Neuhoff, V., Stamm, R. and Eibl, H., *Electrophoresis* 1985, 6, 427-448.
- [31] Neuhoff, V., Arold, N., Taube, D. and Ehrhardt, W., *Electrophoresis* 1988, 9, 255-262.
- [32] Anderson, N. L. and Hickman, B. J., *Anal. Biochem.* 1979, 93, 312-320.
- [33] Sidman, K. E., George, D. E., Barker, W. C. and Hunt, L. T., *Nucl. Acids Res.* 1988, 16, 1869-1871.
- [34] Taylor, J., Anderson, N. L., Coulter, B. P., Scandora, A. E. and Anderson, N. G., in: Radola, B. J. (Ed.), *Electrophoresis '79*, de Gruyter, Berlin 1980, pp. 329-339.
- [35] Taylor, J., Anderson, N. L. and Anderson, N. G., in: Allen, R. C. and Arnaud, P. (Eds.), *Electrophoresis '81*, de Gruyter, Berlin 1981, pp. 383-400.
- [36] Anderson, N. L., Taylor, J., Scandora, A. E., Coulter, B. P. and Anderson, N. G., *Clin. Chem.* 1981, 27, 1807-1820.
- [37] Taylor, J., Anderson, N. L., Scandora, A. E., Jr., Willard, K. E. and Anderson, N. G., *Clin. Chem.* 1982, 28, 861-866.
- [38] Taylor, J., Anderson, N. L. and Anderson, N. G., *Electrophoresis* 1983, 4, 338-345.
- [39] Anderson, N. L. and Taylor, J., in: *Proceedings of the Fourth Annual Conference and Exposition of the National Computer Graphics Association*, Chicago, June 26-30, 1983, pp. 69-76.
- [40] Anderson, N. L., Hofmann, J.-P., Gemmell, A. and Taylor, J., *Clin. Chem.* 1984, 30, 2031-2036.
- [41] Anderson, L., in: Schaefer-Nielsen, C. (Ed.), *Electrophoresis '88*, VCH Verlagsgesellschaft, Weinheim 1988, pp. 313-321.
- [42] Neidhardt, F. C., Appleby, D. A., Sankar, P., Hutton, M. E. and Phillips, T. A., *Electrophoresis* 1989, 10, 116-121.
- [43] Gil, G., Goldstein, J. L., Slaughter, C. A. and Brown, M. S., *J. Biol. Chem.* 1986, 261, 3710-3716.

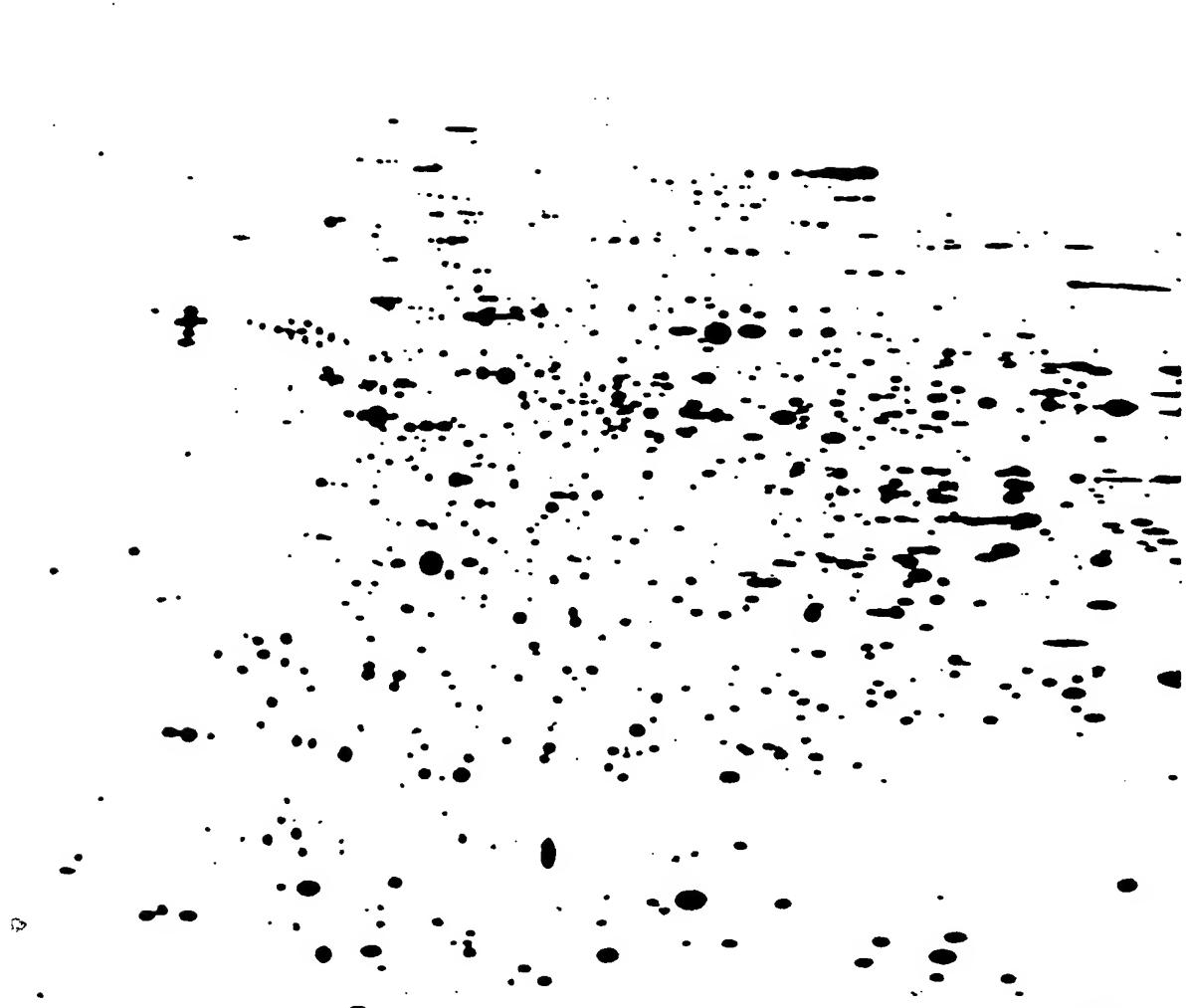
6 Addendum I: Figures 1-13

Figure 1. Synthetic representation of the standard rat liver 2-D master pattern, rendered as a greyscale image using a videoprinter.

Fig. 2. Schem

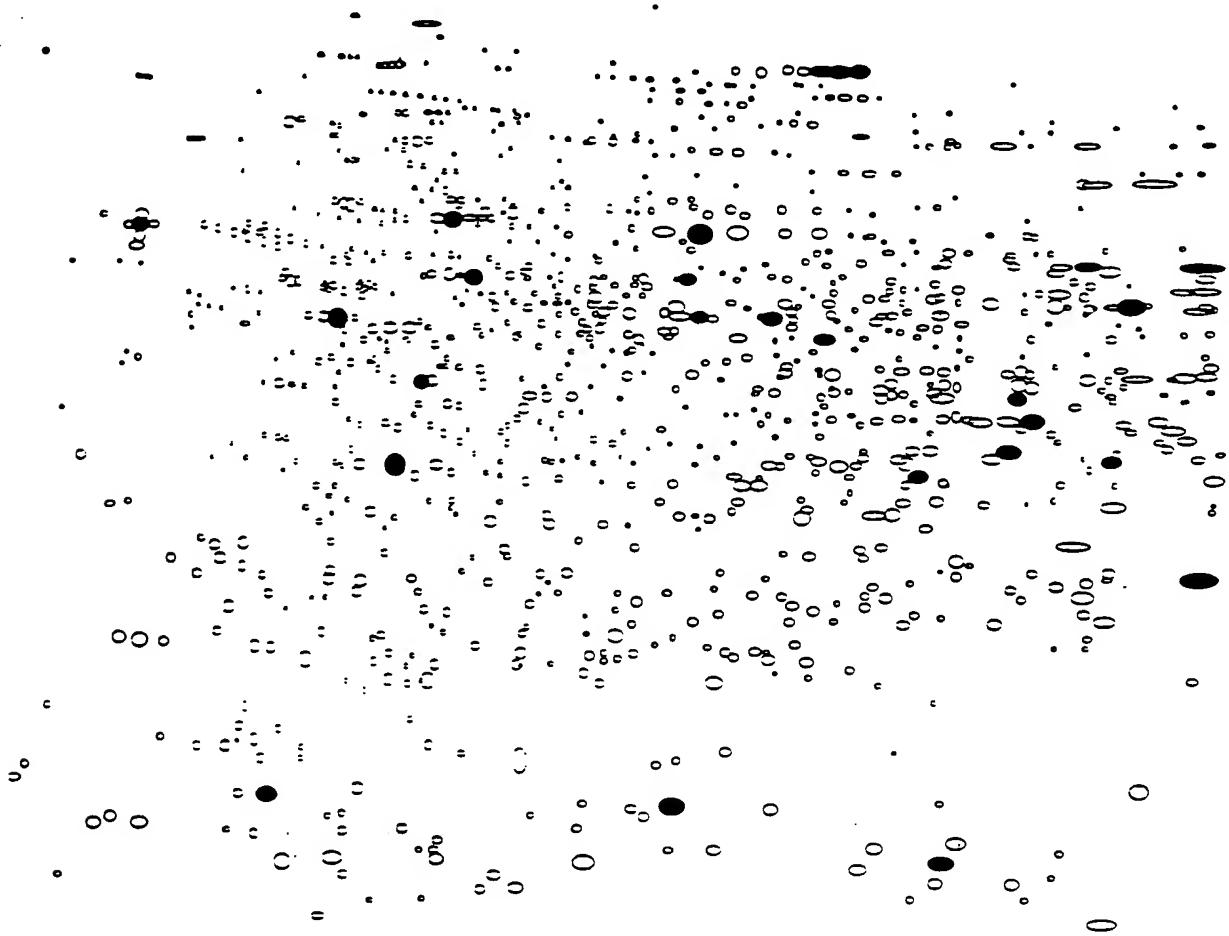


Fig. 2. Schematic representation of the master pattern (the same as Fig. 1), useful as an aid in relating specific areas of Fig. 1 and the following detailed prints.

1

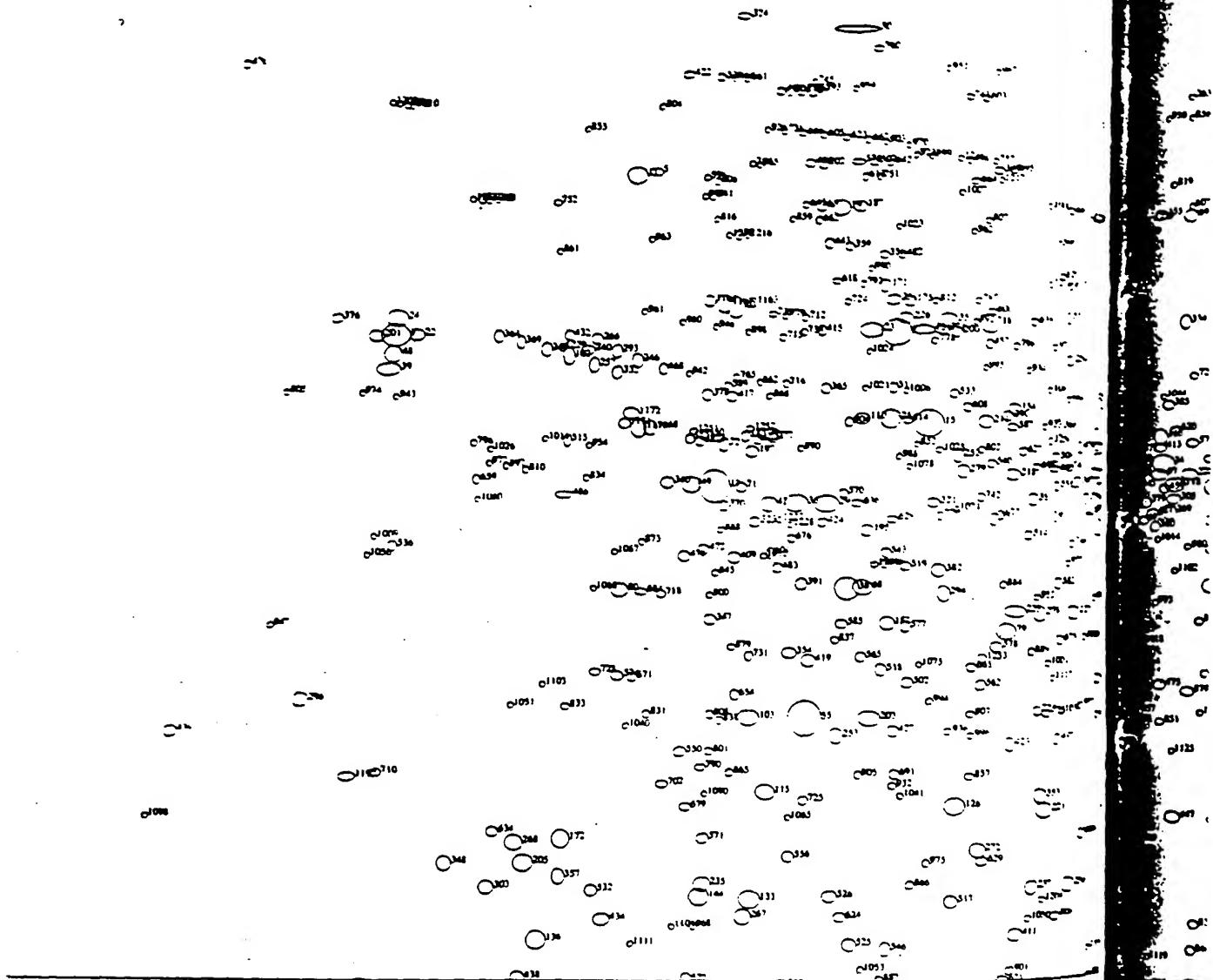


Figure 3. Upper left (high molecular weight, acidic) quadrant (#1) of the rat liver map, showing spot numbers.

2

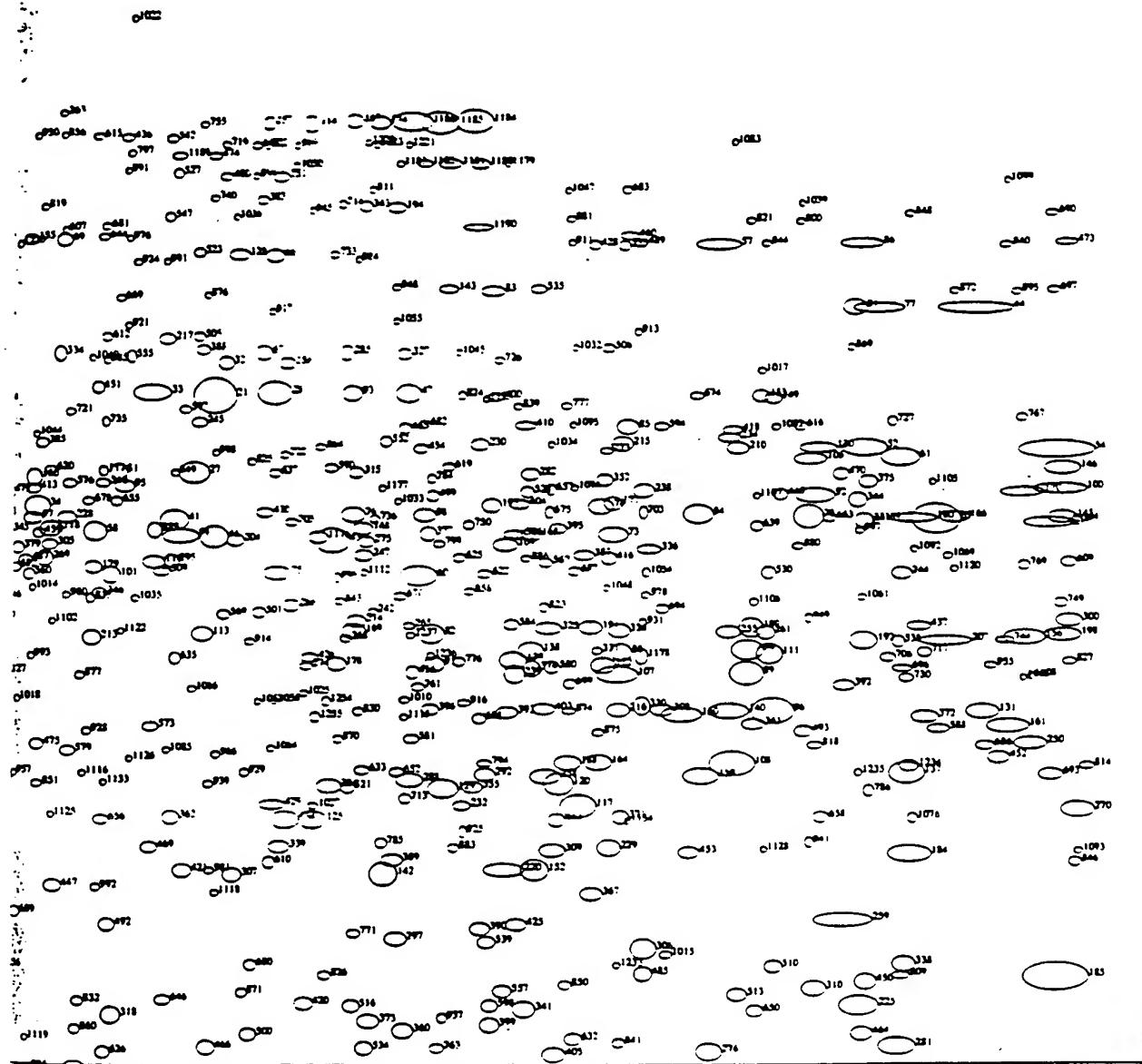
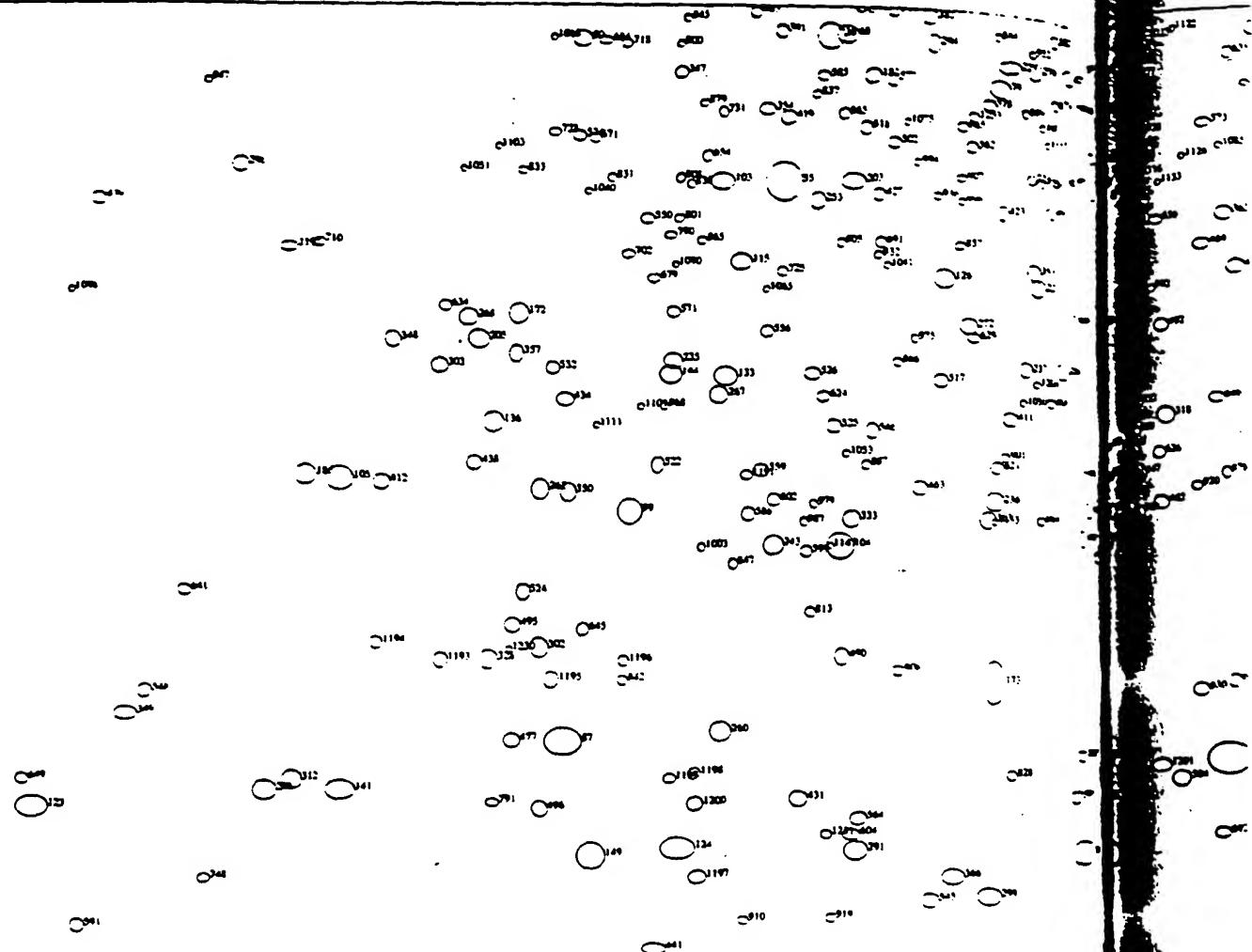


Figure 4. Upper right (high molecular weight, basic) quadrant (#2) of the rat liver map, showing spot numbers.



3

Figure 5. Lower left (low molecular weight, acidic) quadrant (#3) of the rat liver map, showing spot numbers.

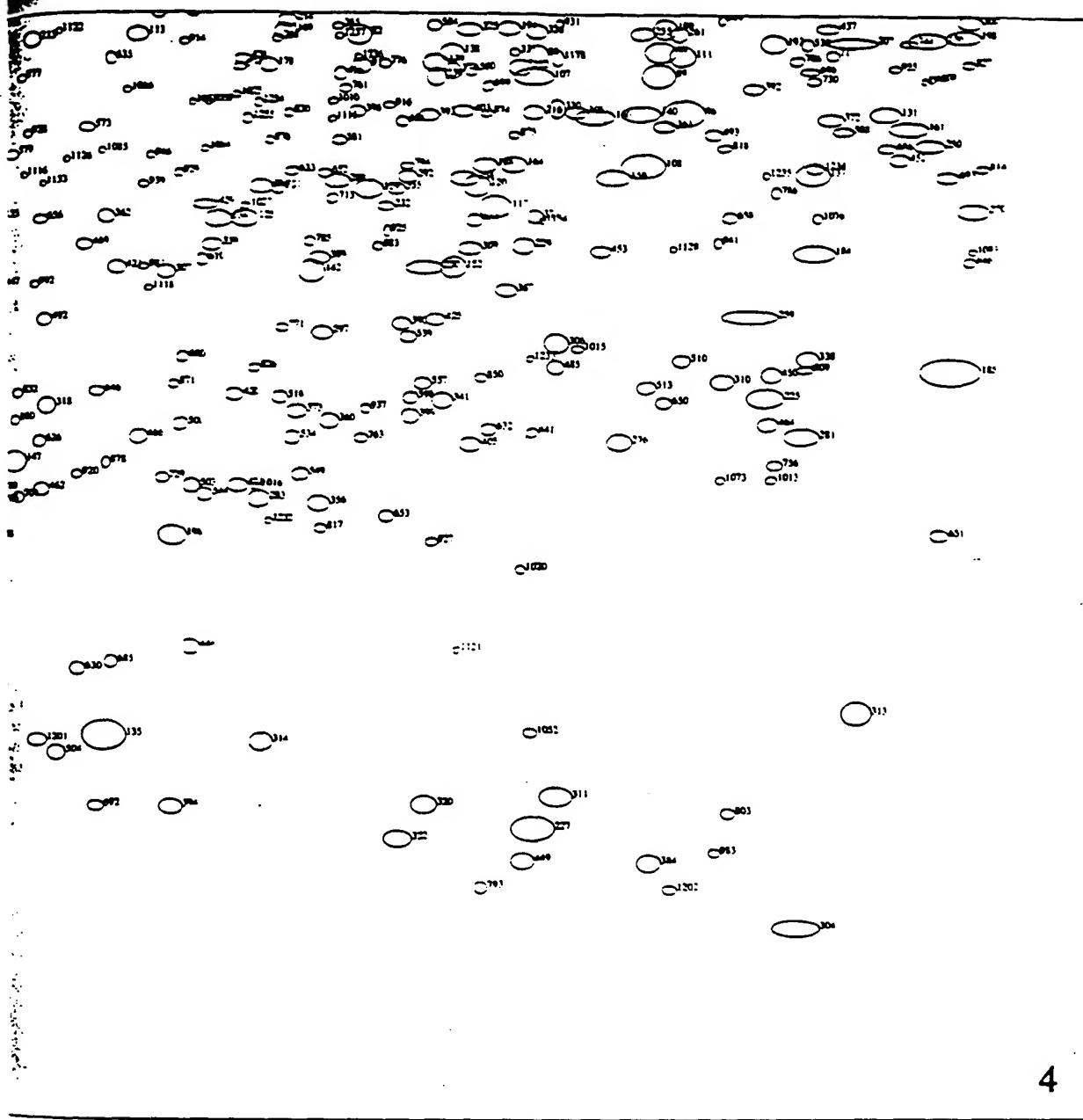


Figure 6. Lower right (low molecular weight, basic) quadrant (#4) of the rat liver map, showing spot numbers.

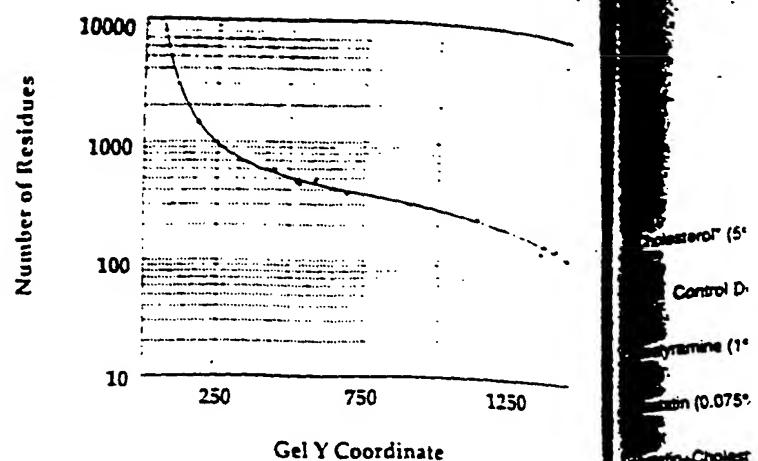
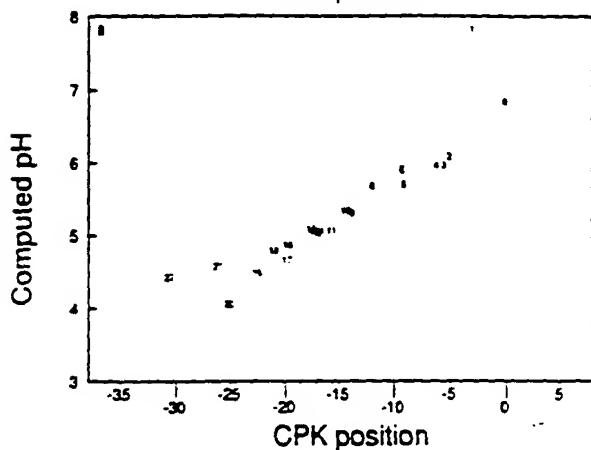
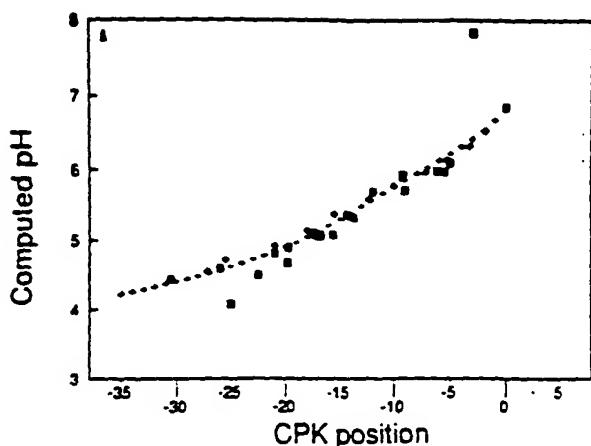


Figure 6. Plot of number of amino acids versus gel λ -position, with fitted curve used to predict molecular mass of unidentified proteins

◀ Figure 7. (a) Plot of computed isoelectric point versus gel λ -position for two sets of carbamylated standard proteins (rabbit muscle CPK [+] and human hemoglobin β chain, filled diamonds) and several other proteins (shaded squares). (b) The identities of the various proteins represented by the squares are indicated by the numbers in corresponding positions on (a); these refer to Table 4.

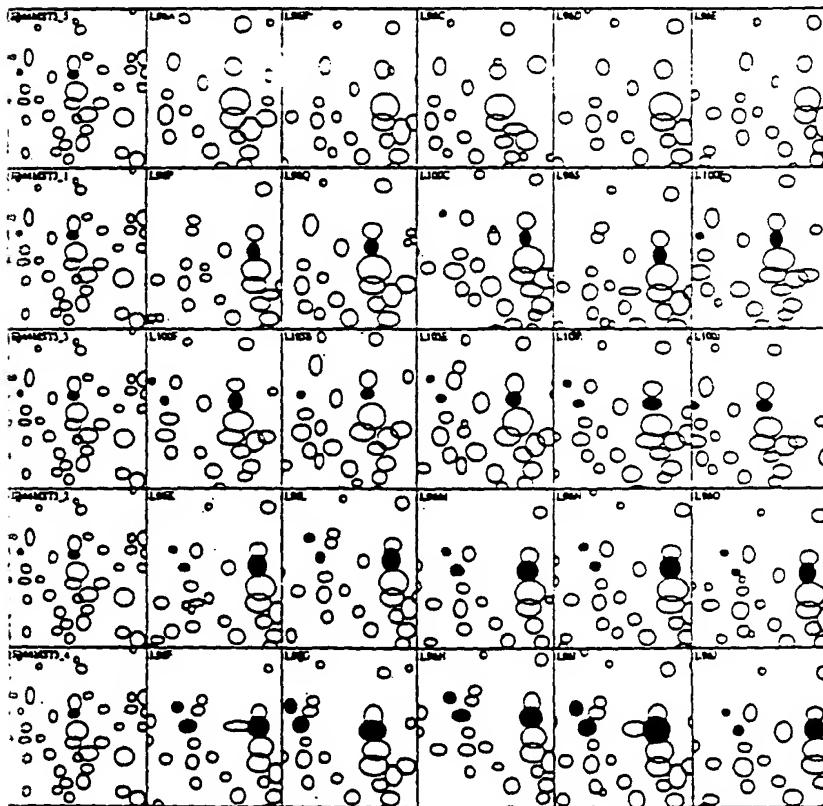


Figure 9. Montage showing effects in the region of MSN:413. The montage shows a small window into one portion of the 2-D pattern, one row of windows for each experimental group, and one panel for each gel in the experiment. The left-most pattern in each row is a group-specific copy of the master pattern followed by the patterns for the five individual rats in the group. The highlighted protein spots (filled circles) are spot 413 (on the right of each panel; identified as cytosolic HMG-CoA reductase) and two modified forms of it (1250 and 933). From the top, the rows (experimental groups) are: high cholesterol, controls, cholestyramine, lovastatin, and lovastatin plus cholestyramine.

Regulation of Rat Liver 413

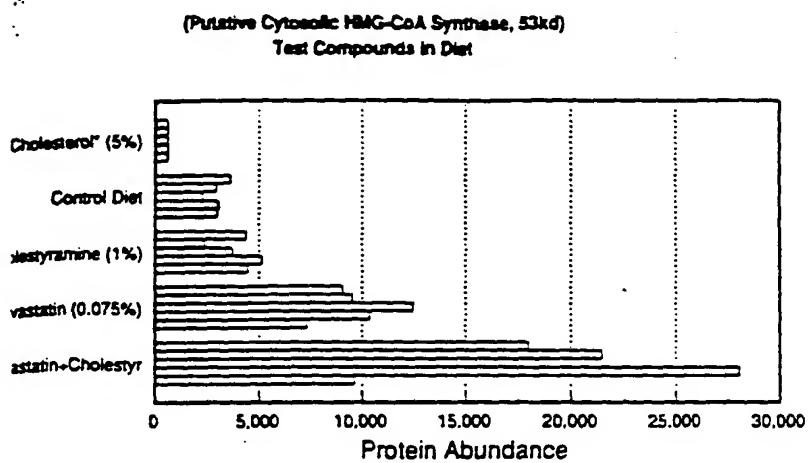


Figure 10. Bargraph showing the quantitative effects of various treatments on the abundance of MSN:413 (cytosolic HMG-CoA synthase) in the gels of Fig. 9.

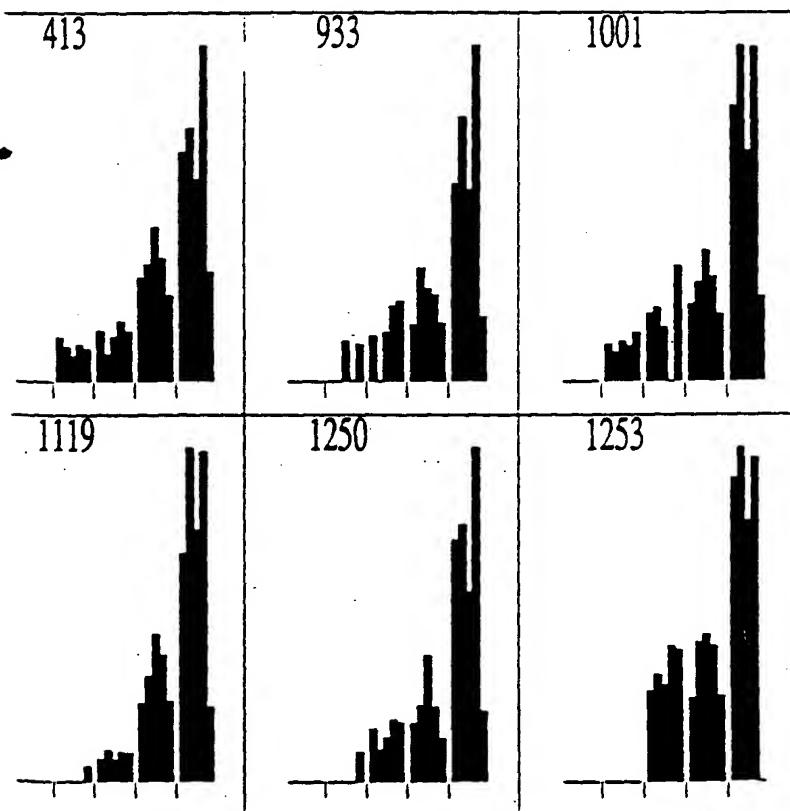


Figure 11. Bargraphs of a series of six coregulated spots including MSN:413. In the bargraphs, the abundances of the appropriate spot (master spot number shown at the top of the panel) in each animal are shown. The five five-animal groups are in the order (left to right): high cholesterol, control, cholestyramine, lovastatin, and lovastatin plus cholestyramine. Each bar within a group represents one experimental animal liver (one 2-D gel). Note the correlated expression of the 6 spots, especially in the two far right (most strongly induced) groups.

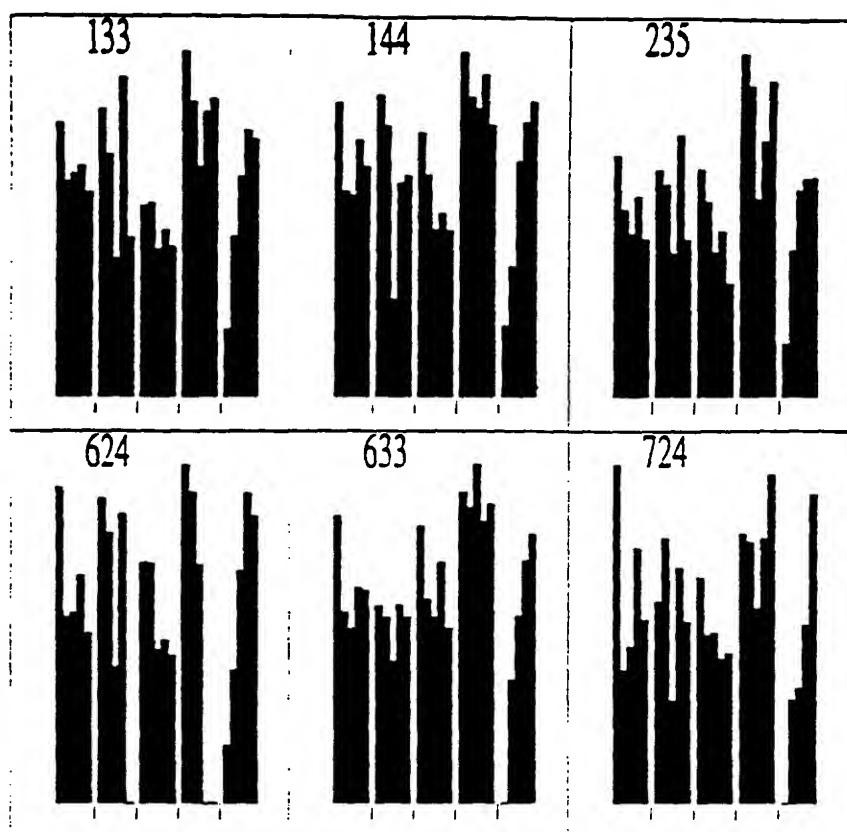


Figure 12. Data on a second coregulated group of spots, presented as in Fig. 11. The fourth experimental group (lovastatin) shows a modest induction, while the fifth group (lovastatin plus cholestyramine) does not.

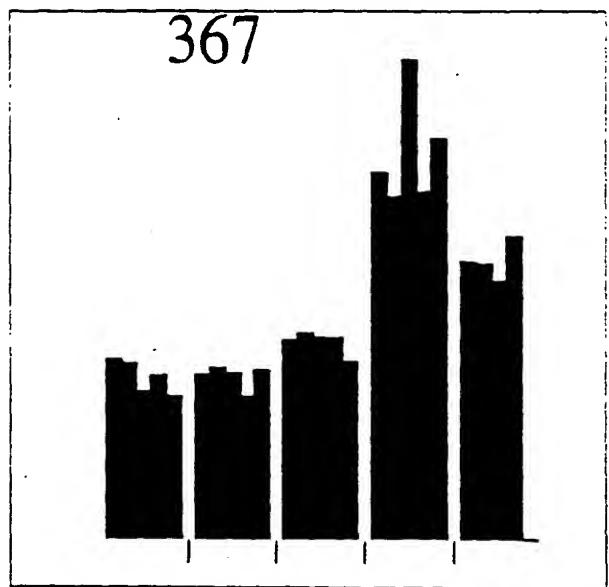


Figure 13. Data on spot MSN:367, presented as in Fig. 11. This protein shows unambiguously the anti-synergistic effect of lovastatin and cholestyramine (fifth group) as compared to lovastatin (fourth group). This response contrasts strongly with the regulation pattern seen in Fig. 11.

Marker
311
589
592
825
833
836
1284
1332
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145

Table 1. Master table of proteins in the rat liver database^a

MSN	X	Y	CPK ₀₁	SDSMW	MSN	X	Y	CPK ₀₁	SDSMW	MSN	X	Y	CPK ₀₁	SDSMW
3	311	434	<-35.0	63,800	85	1119	536	-9.9	53,800	174	1364	183	-6.7	162,800
5	568	263	-24.3	102,800	96	1731	756	-2.0	40,700	175	825	383	-15.7	66,300
8	812	426	-16.0	64,800	97	1033	566	-11.4	51,600	177	1582	553	-3.6	52,600
11	549	268	-25.2	101,000	98	1406	565	-6.1	51,700	178	1321	710	-7.2	43,000
15	845	520	-15.3	55,200	99	578	1149	-23.8	25,000	179	1089	615	-10.4	48,300
17	629	569	-21.6	50,000	100	2004	538	>0.0	53,700	180	1866	567	-0.5	51,600
18	906	414	-14.0	65,300	101	1106	623	-10.1	47,900	181	411	295	-32.1	91,200
19	755	298	-17.5	90,200	102	482	455	-28.5	61,300	182	804	730	-16.2	42,000
20	649	403	-20.9	67,800	103	665	830	-20.2	37,300	184	1860	896	-0.6	34,500
21	1204	448	-8.7	62,100	104	773	1182	-17.0	23,800	185	1997	1017	>0.0	29,800
22	332	434	<-35.0	63,800	105	312	1117	<-35.0	26,100	186	279	1113	<-35.0	26,300
23	787	424	-16.6	65,000	106	1769	509	-1.5	56,100	187	773	296	-17.0	90,800
24	313	417	<-35.0	66,000	107	1585	720	-3.6	42,500	188	1538	807	-4.2	38,400
25	807	516	-16.1	55,500	108	1692	807	-2.4	38,300	191	1560	674	-3.9	44,900
27	1184	524	-9.0	54,800	109	1482	593	-4.8	49,700	192	1818	687	-0.9	44,200
28	1263	446	-8.0	62,400	110	778	516	-16.9	55,500	193	1469	555	-5.0	52,400
29	743	605	-17.8	49,000	111	1728	700	-2.0	43,500	194	1380	266	-6.4	101,600
30	768	112	-17.2	348,600	113	1191	680	-8.9	44,500	195	784	532	-16.7	47,300
32	1216	417	-8.6	66,000	114	1298	185	-7.5	160,800	196	1227	1185	-8.4	23,700
33	1145	445	-9.5	62,500	115	682	907	-19.6	34,100	197	667	553	-20.1	52,600
34	1037	555	-11.3	52,400	116	1146	610	-9.5	48,700	198	2006	681	>0.0	44,500
35	863	412	-14.9	66,600	117	1548	849	-4.1	36,500	199	1711	674	-2.2	44,900
36	712	606	-18.7	48,900	118	1050	577	-11.1	50,800	200	872	424	-14.7	65,000
38	763	694	-17.3	43,800	120	1530	828	-4.3	37,400	201	292	435	<-35.0	63,700
39	304	470	<-35.0	59,800	121	838	423	-15.4	65,200	202	736	253	-18.0	107,800
41	1165	569	-9.2	51,400	122	1572	712	-3.8	42,900	203	786	829	-16.7	37,400
42	684	607	-19.6	48,800	123	23	1433	<-35.0	15,300	204	1224	589	-8.5	50,000
43	1318	589	-7.3	50,000	124	621	1474	-21.9	13,900	205	439	983	-30.9	31,100
44	1924	362	-0.1	74,600	125	1298	862	-7.5	36,000	206	1994	571	>0.0	51,300
46	1203	586	-8.7	50,200	126	872	921	-14.7	33,500	207	1895	687	-0.3	44,200
47	1391	447	-6.3	62,300	127	1000	717	-12.0	42,600	208	240	1418	<-35.0	15,800
48	309	454	<-35.0	61,500	128	1229	311	-8.4	86,100	210	1700	499	-2.3	57,000
49	605	587	-22.5	50,100	129	1422	832	-5.8	37,300	211	902	517	-14.1	55,400
50	621	535	-21.8	53,900	130	1776	490	-1.4	57,000	213	1087	684	-10.4	44,400
51	1113	522	-10.0	55,000	131	1930	757	-0.1	40,700	214	1340	668	-7.0	45,200
52	1820	499	-0.9	57,000	132	660	537	-20.4	53,800	215	1581	495	-3.5	57,300
53	725	177	-18.3	170,800	133	666	1019	-20.2	29,700	216	1585	755	-3.6	40,700
54	2001	500	>0.0	56,800	134	1271	862	-7.9	36,000	217	1159	383	-9.3	69,300
55	722	830	-18.4	37,300	135	1161	1389	-9.3	16,800	218	931	572	-13.5	51,200
56	678	533	-19.6	54,100	136	453	1063	-29.7	28,100	219	713	177	-18.7	170,500
57	1682	302	-2.5	89,000	137	1858	823	-0.6	37,700	220	1479	911	-4.9	33,900
58	1091	580	-10.3	50,600	138	1504	697	-4.6	43,700	221	965	927	-12.8	33,300
59	1171	585	-9.2	50,300	139	1488	707	-4.8	43,200	223	934	716	-13.5	42,700
60	1400	624	-6.2	47,800	140	1689	756	-2.4	40,700	225	1812	1045	-1.0	28,800
61	1853	508	-0.6	56,200	141	311	1417	<-35.0	15,800	226	821	411	-15.8	66,800
62	1888	567	-0.4	51,500	142	1366	915	-6.7	33,800	227	1586	1483	-3.6	13,600
65	735	297	-18.1	90,500	143	1429	346	-5.7	77,900	228	1065	567	-10.8	51,600
66	1263	312	-8.0	85,900	144	615	1017	-22.1	29,800	229	1577	890	-3.7	34,800
67	1252	407	-8.1	67,300	145	2006	566	>0.0	51,600	230	1458	496	-5.2	57,300
68	779	682	-16.8	43,900	146	2006	518	>0.0	55,300	232	1440	849	-5.5	36,500
69	1064	296	-10.8	90,800	147	1070	1108	-10.7	26,500	234	1692	489	-2.4	57,900
71	656	589	-20.6	50,000	148	1347	578	-6.9	50,800	235	618	1004	-22.0	30,300
72	638	545	-21.2	53,100	149	541	1481	-25.7	13,700	236	920	1138	-13.7	25,400
73	1582	583	-3.6	50,400	150	1645	760	-2.8	40,500	237	952	1008	-13.1	30,200
74	1570	556	-3.8	52,300	151	1269	236	-7.9	117,000	238	1611	541	-3.2	53,500
75	1264	621	-8.0	48,000	152	1507	911	-4.5	33,900	239	1489	720	-4.8	42,500
76	1338	564	-7.0	51,800	153	1722	448	-2.1	62,100	240	501	448	-27.7	62,100
77	1833	363	-0.8	74,400	154	932	503	-13.5	56,600	241	1820	569	-0.9	51,400
78	1767	565	-1.5	51,700	155	1001	294	-11.4	91,400	242	1357	658	-6.8	45,800
79	925	738	-13.6	41,600	156	1970	684	>0.0	44,400	243	711	1182	-18.7	23,800
80	534	698	-26.1	43,600	157	1258	183	-8.1	162,400	244	1855	621	-0.6	48,000
81	1811	363	-1.0	74,500	158	1275	417	-7.8	65,900	245	1189	474	-8.9	59,300
82	1412	681	-6.0	44,500	159	1663	820	-2.6	37,800	246	551	459	-25.1	61,000
83	1471	347	-5.0	77,500	160	1034	527	-11.4	54,600	247	1348	604	-6.9	49,100
84	1662	563	-2.7	51,800	161	1953	771	>0.0	40,000	248	460	448	-29.3	62,100
85	1596	479	-3.4	58,900	162	1020	1482	-11.6	13,700	249	1733	451	-1.9	61,800
86	1817	301	-0.9	89,100	164	1566	806	-3.8	38,400	250	1974	788	>0.0	39,200
87	516	1371	-27.0	17,400	165	1905	565	-0.2	51,700	251	808	392	-16.1	69,500
88	1589	698	-3.5	43,600	167	1340	181	-7.0	164,900	252	874	553	-14.6	52,500
89	1706	719	-2.2	42,500	168	1506	583	-4.6	50,400	253	753	848	-17.6	36,500
90	651	329	-20.8	81,700	169	1338	678	-7.0	44,700	254	995	450	-12.1	61,900
91	1415	710	-6.0	43,000	170	1969	541	>0.0	53,500	255	1690	679	-2.4	44,600
92	1773	545	-1.4	53,200	171	800	378	-16.3	71,800	256	994	1006	-12.1	30,200
93	1338	446	-7.0	62,300	172	476	958	-28.7	32,100	257	508	464	-27.4	60,400
94	1708	696	-2.2	43,700	173	919	1314	-13.7	19,300	258	1517	820	-4.4	37,800

Master table of proteins in the rat liver database, showing spot master number, gel position (x and y), isoelectric point relative to CPK standards, and predicted molecular mass (from the standard curve of Fig. 8).

MSN	X	Y	CPK ₁	SDSMW	MSN	X	Y	CPK ₁	SDSMW	MSN	X	Y	CPK ₁	SDSMW
259	1706	961	-1.1	31,900	345	1006	578	-11.9	50,800	426	1296	704	-7.6	43,300
260	661	1361	-20.4	17,700	346	1095	640	-10.3	46,800	427	810	843	-16.0	36,800
261	1725	679	-2.0	44,600	347	625	728	-21.7	42,000	428	1565	303	-3.9	88,700
262	486	1127	-28.0	25,800	348	361	963	-35.3	31,100	429	1250	847	-8.0	36,800
263	1053	172	-10.9	177,400	349	110	1343	<-35.0	18,300	430	1253	562	-8.1	51,800
265	1380	673	-6.3	45,000	350	521	1130	-26.7	25,700	431	734	1426	-18.1	15,500
266	510	437	-27.3	63,400	351	912	619	-13.9	48,100	432	483	433	-29.5	63,800
267	660	1038	-20.4	29,000	352	1574	530	-3.7	54,300	434	518	1041	-26.9	28,800
268	430	951	-31.0	31,800	353	951	912	-12.9	33,900	435	1020	1170	-11.6	24,300
269	1044	606	-11.2	48,900	354	706	762	-18.9	40,400	436	1122	196	-9.8	147,800
270	2019	853	>0.0	36,300	355	1450	830	-5.3	37,300	437	1870	673	-0.5	45,000
271	857	422	-15.0	65,200	356	1374	1152	-6.5	24,900	438	435	1102	-31.0	26,700
272	895	858	-14.2	31,700	357	474	997	-28.7	30,600	439	86	847	<-35.0	36,800
274	1282	712	-7.6	42,900	358	798	346	-16.3	77,800	440	1740	544	-1.8	53,200
275	1350	590	-6.9	49,900	359	764	338	-17.3	79,400	441	599	1571	-22.6	10,800
276	1670	1069	-2.6	27,100	360	1384	1068	-6.4	27,900	443	743	335	-17.8	80,100
277	688	538	-19.4	53,700	361	1713	769	-2.1	40,100	445	801	668	-16.2	45,200
278	961	718	-13.0	42,600	362	1161	859	-9.3	36,100	447	1050	826	-11.1	33,300
279	879	570	-14.5	51,300	363	914	1156	-13.8	24,800	448	1245	1298	-8.2	19,800
281	1848	1084	-0.7	27,300	364	412	435	-32.0	63,700	449	1576	1516	-3.7	12,800
282	1505	525	-4.6	54,800	365	741	486	-17.9	58,200	450	1818	1021	-0.9	29,800
283	1313	1147	-7.3	25,100	366	878	1503	-14.6	13,000	451	1094	440	-10.3	63,100
284	1314	829	-7.3	37,400	367	1560	935	-3.9	33,000	452	1945	802	>0.0	38,800
285	1332	408	-7.1	67,200	368	963	520	-12.4	55,200	453	1652	894	-2.8	34,600
286	1277	652	-7.8	46,100	369	434	441	-31.0	63,000	454	1403	500	-6.1	56,800
288	1391	824	-6.3	37,600	370	639	610	-21.2	48,700	455	1394	718	-6.3	42,800
289	1147	579	-9.5	50,700	371	1587	860	-3.6	36,100	457	905	436	-14.0	63,500
290	925	511	-13.6	55,900	372	1875	762	-0.5	40,400	459	1038	581	-11.3	50,500
291	787	1476	-16.6	13,900	373	1351	1059	-6.8	28,300	460	1598	294	-3.4	91,400
292	1462	818	-5.1	37,800	374	1506	715	-4.6	42,700	461	1526	863	-4.3	35,800
293	531	449	-26.3	62,000	375	1823	532	-0.9	54,200	462	1098	1137	-10.2	25,400
294	860	696	-14.9	43,600	376	254	417	<-35.0	65,900	463	849	1125	-15.2	25,800
295	1162	609	-9.3	48,700	377	1409	583	-6.1	50,400	464	1814	1072	-0.9	27,800
296	218	814	<-35.0	38,000	378	621	494	-21.8	57,500	465	1388	481	-6.3	58,700
297	1377	979	-6.5	31,300	379	1017	595	-11.7	49,600	466	1194	1064	-8.9	27,300
299	913	1523	-13.9	12,400	381	953	598	-13.1	49,400	468	577	467	-23.9	60,100
300	2012	667	>0.0	45,300	382	856	674	-15.0	44,900	469	1140	888	-9.6	34,900
301	702	178	-19.0	168,200	383	1252	258	-8.1	105,300	470	1797	524	-1.1	54,800
302	494	1280	-28.1	20,400	384	1699	1518	-2.3	12,500	471	1293	1133	-7.6	25,500
303	403	1008	-32.6	30,100	385	1042	493	-11.2	57,500	472	618	655	-21.9	46,000
304	1843	1585	-0.7	10,300	386	1490	583	-4.7	50,400	473	2009	299	>0.0	89,800
305	1049	503	-11.1	49,800	387	1554	603	-4.0	49,100	474	1205	215	-8.7	131,300
306	1608	989	-3.3	30,900	388	1193	404	-8.9	67,700	475	1035	788	-11.4	39,200
307	1219	916	-8.5	33,700	389	1374	902	-6.5	34,300	476	160	155	<-35.0	207,600
308	1527	755	-3.0	40,700	390	1456	969	-5.2	31,700	477	469	1370	-28.9	17,400
309	1524	892	-4.4	34,700	391	718	690	-18.5	44,000	478	599	662	-22.8	45,600
310	1769	1028	-1.5	29,400	392	1799	732	-1.1	41,900	479	1009	540	-11.8	53,900
311	1609	1451	-3.3	14,700	393	1482	758	-4.8	40,600	480	1216	235	-8.6	117,400
312	266	1408	<-35.0	16,100	394	1227	1461	-8.4	14,400	482	816	346	-15.9	77,800
313	1902	1365	-0.3	17,600	395	1530	577	-4.3	50,800	483	603	673	-19.3	44,900
314	1316	1395	-7.3	16,600	396	1410	755	-6.0	40,800	485	1608	1013	-3.3	30,000
315	1341	523	-7.0	54,900	397	912	256	-13.9	106,400	486	478	599	-28.6	49,300
318	1104	1053	-10.1	28,500	399	1465	1063	-5.0	28,100	487	1025	607	-11.5	48,600
320	1480	1459	-4.9	14,400	400	1473	450	-4.9	61,900	488	1045	1186	-11.2	23,700
321	650	603	-15.1	49,100	401	1029	1140	-11.5	25,300	489	1609	301	-3.3	89,200
322	1454	1494	-5.3	13,300	403	1516	754	-4.4	40,800	490	775	1289	-17.0	20,100
323	670	626	-20.0	47,700	404	1495	554	-4.7	52,500	491	692	178	-19.3	169,300
324	655	101	-20.6	420,500	405	1525	1092	-4.3	27,100	492	1100	964	-10.2	31,800
325	1521	675	-4.4	44,800	406	723	252	-18.4	106,000	493	1760	776	-1.6	39,700
326	1587	677	-3.6	44,700	409	650	663	-20.8	45,500	494	882	247	-14.5	110,700
327	1368	408	-6.3	67,000	410	1501	478	-4.6	59,000	495	470	1258	-28.9	21,200
328	448	1291	-30.0	20,100	411	936	1057	-13.4	28,300	496	494	1436	-28.1	15,200
330	1608	751	-3.3	40,900	412	350	1120	-35.9	26,000	497	980	852	-12.5	36,600
331	1566	697	-3.8	43,700	413	1033	538	-11.4	53,700	499	1414	546	-6.0	53,100
332	531	471	-26.3	59,800	415	737	425	-18.0	64,900	500	1234	1072	-8.3	27,800
333	784	1156	-16.7	24,700	416	1578	606	-3.7	48,900	501	1246	659	-8.2	45,700
334	1050	407	-10.9	67,300	417	646	496	-21.0	57,300	502	824	792	-15.7	39,000
335	1593	303	-3.5	88,500	418	1695	482	-2.3	58,600	503	1246	1134	-8.2	25,500
336	1616	598	-3.2	49,400	419	725	770	-18.3	40,000	504	1115	1407	-9.9	16,200
338	1854	1004	-0.6	30,300	420	1289	1041	-7.7	28,900	505	1189	391	-8.9	68,000
339	1265	888	-8.0	34,900	421	1171	912	-9.1	33,900	506	1578	402	-3.7	102,000
340	501	585	-23.6	50,300	422	599	162	-22.8	193,700	507	787	250	-16.6	102,000
341	1497	1047	-4.7	28,700	423	929	856	-13.6	36,200	508	979	552	-12.5	52,000
343	1351	265	-6.8	102,200	424	739	625	-17.9	47,700	509	1153	619	-9.4	48,700
344	1813	549	-0.9	52,800	425	1490	965	-4.7	31,800	510	1730	1006	-2.0	30,200

ISN	X	Y	CPK _{d1}	SDSMW	ISN	X	Y	CPK _{d1}	SDSMW	ISN	X	Y	CPK _{d1}	SDSMW
511	809	484	-16.0	58,400	506	619	269	-21.9	100,500	674	1651	448	-2.7	62,100
512	1099	533	-10.2	54,100	507	1176	461	-9.1	60,700	675	1523	562	-4.4	51,800
513	1696	1034	-2.3	29,200	508	1465	1044	-5.0	28,800	676	708	642	-18.8	46,700
514	948	636	-13.2	47,100	509	741	1188	-17.9	23,600	677	919	615	-13.7	48,300
515	481	543	-28.5	53,400	600	907	402	-14.0	68,000	678	1085	551	-10.5	52,700
516	1334	1044	-7.1	28,800	601	687	658	-19.5	45,800	679	600	823	-22.7	33,400
517	868	1021	-14.8	29,700	602	712	1138	-18.7	25,400	680	1237	1004	-8.3	30,300
518	798	779	-16.3	39,600	603	898	181	-14.1	165,200	681	1103	283	-10.1	95,100
519	822	670	-15.7	45,100	604	783	1461	-16.7	14,400	682	1406	477	-6.1	59,100
520	632	165	-21.5	189,000	605	736	223	-18.0	125,300	683	1596	249	-3.4	109,800
521	1332	830	-7.1	37,300	606	629	273	-21.6	98,700	684	555	699	-24.8	43,500
522	603	1104	-22.6	26,800	607	1064	286	-10.8	94,000	685	1167	1313	-9.2	19,300
523	1190	309	-8.9	86,800	608	583	503	-14.5	56,700	686	1932	790	0.0	39,100
524	479	1226	-28.6	22,300	609	2012	610	>0.0	48,700	687	1545	619	-4.1	48,100
525	768	1066	-17.2	28,000	610	1255	903	-8.1	34,200	688	1456	764	-5.2	40,300
526	747	1016	-17.7	29,800	612	1103	391	-10.1	69,600	689	1011	953	-11.8	32,300
527	1170	231	-9.2	119,600	613	778	265	-16.9	102,000	690	1995	270	>0.0	100,200
528	1502	542	-4.6	53,400	614	824	518	-15.7	55,400	691	812	888	-16.0	34,900
529	1728	620	-2.0	48,000	615	1095	195	-10.3	149,100	692	1154	1461	-9.4	14,400
530	507	1011	-27.4	30,000	616	1759	478	-1.6	58,000	693	1993	819	>0.0	37,800
531	870	489	-14.7	57,900	617	994	372	-12.1	72,900	694	1628	656	-3.0	45,900
532	1347	1085	-6.9	27,300	618	751	374	-17.6	72,400	695	928	254	-13.6	107,000
533	1513	346	-4.5	77,800	619	1429	518	-5.7	55,300	696	1854	715	-0.6	42,700
534	306	654	<-35.0	46,000	620	1050	520	-11.1	55,200	697	1997	345	>0.0	78,000
535	1851	689	-0.7	44,100	621	923	1105	-13.7	26,600	698	957	563	-13.0	51,800
536	1463	982	-5.1	31,100	622	1462	622	-5.1	47,900	699	1540	730	-4.2	42,000
537	909	561	-13.9	52,000	623	759	225	-17.4	124,000	702	577	900	-23.8	34,400
538	625	289	-21.7	93,100	624	758	1038	-17.4	29,000	703	1610	562	-3.2	51,900
539	1164	198	-9.2	146,200	625	1438	606	-5.5	48,800	705	1278	571	-7.8	51,200
540	803	655	-16.2	45,900	626	1096	1089	-10.2	27,200	706	1841	704	-0.7	43,300
541	1259	1143	-8.0	25,200	627	942	548	-13.3	53,000	707	1018	1386	-11.7	16,900
542	856	1526	-15.0	12,200	628	809	621	-16.0	48,000	709	1074	1145	-10.7	25,100
543	803	1071	-16.2	27,800	629	899	979	-14.1	31,300	710	293	889	<-35.0	34,800
544	1162	274	-9.3	98,400	630	1135	1321	-9.6	19,100	712	720	412	-18.5	66,600
545	128	1321	<-35.0	19,000	631	979	615	-12.5	48,300	713	1386	841	-6.4	36,800
546	1355	1122	-6.8	25,900	632	1542	1076	-4.1	27,600	714	1328	263	-7.1	103,100
547	595	866	<-23.0	35,800	633	1345	814	-6.9	38,000	715	698	433	-19.1	63,900
548	1369	494	-6.6	57,500	634	409	950	-32.2	32,400	716	701	481	-19.0	58,700
549	992	405	-12.2	67,600	635	1165	704	-9.2	43,300	717	1875	699	-0.5	43,600
550	1125	410	-9.8	66,900	636	774	604	-17.0	49,000	718	575	702	-23.9	43,400
551	705	975	-18.9	31,400	637	1263	524	-8.0	54,800	719	1216	204	-8.6	140,400
552	1477	1030	-4.9	29,300	638	952	411	-13.1	66,700	721	1069	464	-10.8	60,400
553	980	583	-12.5	50,400	639	1717	575	-2.1	51,000	722	1272	506	-7.9	56,400
554	700	1109	-19.1	26,400	640	994	292	-12.1	92,000	723	958	822	-13.0	37,700
555	1028	621	-11.5	48,000	641	165	1224	<-35.0	22,400	724	763	395	-17.3	69,100
556	896	794	-14.1	38,900	642	803	251	-16.2	108,900	725	720	916	-18.5	33,700
557	789	1446	-16.6	14,900	643	719	296	-18.5	90,700	726	1476	415	-4.9	66,200
558	777	766	-16.9	40,200	644	1100	294	-10.2	91,400	727	1846	473	-0.7	59,400
559	980	328	-12.5	81,900	645	534	1263	-26.1	21,000	728	510	783	-27.3	39,400
560	1519	611	-4.4	48,600	646	1153	1038	-9.4	29,000	729	1217	1126	-8.6	25,800
561	1212	661	-8.6	45,800	648	1246	204	-8.2	140,000	730	1858	724	-0.6	42,300
562	760	594	-17.4	49,700	649	14	1406	<-35.0	16,200	731	665	765	-20.2	40,300
563	618	956	-21.9	32,100	650	1713	1049	-2.1	28,600	733	1321	312	-7.2	85,900
564	1142	771	-9.6	40,000	651	1986	1183	>0.0	23,800	734	719	427	-18.5	64,600
565	532	787	-26.2	39,300	652	1378	816	-6.5	38,000	735	1101	473	-10.2	59,500
566	771	250	-17.1	109,200	653	1442	1165	-5.5	24,400	736	1359	569	-6.7	51,400
567	1068	534	-10.8	54,100	654	650	806	-20.8	38,400	738	696	220	-19.2	127,600
568	822	734	-15.7	41,800	655	1111	551	-10.0	52,700	739	687	409	-19.5	67,000
569	914	754	-13.8	40,800	656	1095	861	-10.3	36,000	740	1205	256	-8.7	106,200
570	1064	794	-10.8	38,900	657	1524	540	-4.4	53,600	741	995	563	-12.1	51,900
571	1524	714	-4.4	42,800	658	1777	860	-1.4	36,000	742	898	596	-14.1	49,500
572	1392	783	-6.3	39,400	659	391	584	-33.4	50,400	743	881	181	-14.5	165,900
573	982	686	-12.4	44,200	660	977	565	-12.5	51,700	744	1951	686	>0.0	44,200
574	1487	672	-4.8	45,000	661	658	166	-20.5	187,500	745	726	168	-18.3	183,600
575	758	731	-17.4	41,900	662	732	312	-18.1	86,100	746	999	643	-12.0	46,600
576	687	1152	-19.5	24,900	663	1787	567	-1.2	51,500	748	182	1503	<-35.0	13,000
577	930	523	-13.5	55,000	664	888	268	-14.4	100,800	749	2005	649	>0.0	46,300
578	1888	774	-0.4	39,900	665	889	775	-14.3	39,800	750	1448	575	-5.4	51,000
579	642	485	-21.1	58,300	666	715	221	-18.6	126,300	751	792	266	-16.5	101,900
580	1317	519	-7.3	55,300	667	781	227	-16.8	122,400	752	469	296	-28.9	90,600
581	65	1548	<-35.0	11,500	668	646	165	-21.0	189,100	754	664	254	-20.3	107,000
582	1014	614	-11.7	48,400	669	1116	353	-9.9	76,300	755	1195	184	-8.8	161,000
583	732	176	-18.1	172,300	670	1382	643	-6.4	46,600	756	1821	1113	-0.9	26,300
584	1627	478	-3.0	59,000	671	547	789	-25.3	39,200	757	909	246	-13.9	111,000
585	1009	1426	-11.8	15,500	673	984	746	-12.4	41,200	760	790	133	-16.5	264,900

MSN	X	Y	CPKdI	SDSMW	MSN	X	Y	CPKdI	SDSMW	MSN	X	Y	CPKdI	SDSMW
761	1399	733	-6.2	41,800	848	1863	271	-6	99,500	939	1197	827	-8.8	37,500
763	1416	1085	-5.9	27,300	849	1166	523	-9.2	54,900	941	1765	885	-1.5	35,000
764	2020	560	>0.0	51,400	850	1535	1024	-4.2	29,600	942	602	472	-22.7	59,600
765	651	475	-20.8	59,300	851	1035	826	-11.4	37,500	943	312	498	<-35.0	57,100
766	1052	1149	-11.1	25,000	852	834	542	-15.5	53,400	944	993	491	-12.1	57,700
767	1968	468	>0.0	59,800	855	499	220	-27.8	127,100	945	1300	269	-7.5	100,300
768	1330	685	-7.1	44,300	856	1063	194	-10.9	150,500	946	630	423	-21.6	65,100
769	1970	613	>0.0	48,500	857	887	890	-14.4	34,800	947	187	736	<-35.0	41,600
770	857	617	-15.0	48,200	858	1448	639	-5.4	46,900	948	1380	344	-6.5	78,200
771	1337	974	-7.0	31,500	859	706	311	-18.9	86,200	949	1766	665	-1.5	45,400
773	1576	502	-3.7	56,700	860	1070	1066	-10.7	28,000	950	1038	193	-11.3	151,000
775	966	824	-12.8	37,600	861	472	347	-28.8	77,600	951	860	152	-14.9	213,000
776	1438	708	-5.5	43,100	862	674	480	-19.9	58,800	952	957	701	-13.0	43,400
777	1539	458	-4.2	61,000	864	1307	499	-7.4	57,000	954	503	547	-27.6	53,000
778	850	434	-15.1	63,800	865	645	887	-21.0	34,900	955	1938	712	>0.0	42,900
779	700	411	-19.1	66,800	866	827	1004	-15.6	30,300	957	1010	816	-11.8	37,900
780	1052	1136	-11.1	25,500	868	685	494	-19.5	57,400	959	768	174	-17.2	174,900
784	1413	529	-6.0	54,400	869	1807	402	-1.0	68,000	960	596	419	-23.0	65,700
785	1364	885	-6.7	35,000	870	1323	783	-7.2	39,400	961	557	409	-24.8	67,100
786	1822	835	-0.9	37,100	871	1228	1031	-8.4	29,300	962	887	320	-14.4	83,900
787	893	392	-14.3	69,500	872	1904	346	-0.3	77,700	963	564	334	-24.5	80,500
788	616	882	-22.0	35,100	873	556	647	-24.8	46,400	964	969	1155	-12.8	24,800
791	451	1429	-29.8	15,400	874	1540	756	-4.2	40,700	965	671	255	-20.0	106,600
792	777	377	-16.9	72,000	875	1566	777	-3.8	39,700	966	1204	798	-8.7	38,700
793	1536	1543	-4.2	11,700	876	1198	351	-8.8	76,800	967	910	154	-13.9	210,300
794	1461	807	-5.1	38,300	877	1076	720	-10.6	42,500	968	609	1048	-22.3	28,700
796	388	546	-33.6	53,100	878	1161	1111	-9.3	26,400	969	1285	206	-7.7	138,900
797	1126	212	-9.8	133,700	879	647	757	-20.9	40,700	970	822	232	-15.8	119,300
798	833	437	-13.5	63,400	880	1756	594	-1.6	49,700	971	976	437	-12.6	63,400
799	1420	593	-5.9	49,800	881	1543	278	-4.1	97,100	972	403	567	-32.6	51,600
800	1759	279	-1.6	96,500	883	1432	890	-5.7	34,800	974	279	495	<-35.0	57,400
801	624	865	-21.7	35,800	884	922	689	-13.7	44,100	975	844	981	-15.3	31,200
802	896	547	-14.2	53,000	885	1103	414	-10.1	66,400	976	1124	295	-9.8	91,100
803	1775	1468	-1.4	14,200	886	1501	607	-4.6	48,900	977	994	664	-12.1	45,400
804	573	196	-24.0	148,400	887	798	1103	-16.3	26,600	978	1612	642	-3.2	46,700
805	203	494	<-35.0	57,400	888	636	634	-21.3	47,200	979	749	1141	-17.7	25,300
806	980	1039	-12.5	29,000	889	951	759	-13.1	40,600	980	1064	642	-10.8	46,700
807	902	308	-14.1	87,200	890	717	548	-18.6	52,900	981	1197	911	-8.8	33,900
808	625	827	-21.7	37,500	891	1123	229	-9.8	121,200	983	1762	1508	-1.6	12,800
809	1851	1015	-0.7	29,900	892	891	413	-14.3	66,400	984	1344	317	-6.9	84,700
810	440	573	-30.9	51,100	894	1245	234	-8.2	117,800	985	1024	1105	-11.5	26,600
811	1358	249	-6.8	109,700	895	1962	346	>0.0	77,700	987	739	1159	-17.9	24,600
812	851	393	-15.1	69,400	896	1322	626	-7.2	47,700	988	816	555	-15.9	52,400
813	745	1246	-17.8	21,600	897	420	570	-31.4	51,300	990	785	361	-16.7	74,900
814	2028	810	>0.0	38,200	898	662	428	-20.3	64,500	991	1159	317	-9.3	84,500
815	1066	645	-10.4	46,500	899	845	243	-15.3	113,000	992	1090	928	-10.4	33,300
816	629	313	-21.6	85,700	900	624	703	-21.7	43,400	993	1030	701	-11.5	43,400
817	1376	1177	-6.5	24,000	901	931	1094	-13.5	27,000	994	847	811	-15.2	38,200
818	1771	790	-1.4	39,100	903	799	229	-16.3	121,000	995	902	461	-14.1	60,700
819	1045	263	-11.2	103,100	904	765	520	-17.2	55,200	996	888	847	-14.4	36,600
820	984	362	-12.4	74,600	905	775	889	-17.0	34,800	997	1815	579	-0.9	50,700
821	1712	279	-2.2	96,700	907	888	824	-14.4	37,600	998	1205	504	-8.7	56,500
822	1256	205	-8.1	139,200	908	828	1303	-15.6	19,700	999	617	289	-22.0	93,100
823	1517	654	-4.4	46,000	910	681	1544	-19.7	11,700	1000	968	290	-12.8	92,700
824	1442	449	-5.5	62,000	911	1544	301	-4.1	89,100	1001	970	771	-12.7	40,000
825	1240	513	-8.3	55,800	913	1606	387	-3.3	70,400	1002	1736	478	-1.9	58,900
826	1309	1014	-7.4	29,900	914	1237	688	-8.3	44,100	1003	643	1184	-21.1	23,700
827	2012	708	>0.0	43,100	916	1442	749	-5.5	41,100	1006	822	487	-15.8	58,100
828	937	1405	-13.4	16,200	917	1260	367	-8.0	73,700	1007	875	279	-14.6	96,400
830	1342	756	-7.0	40,700	919	764	1541	-17.3	11,700	1009	291	644	<-35.0	46,600
831	562	826	-24.5	37,500	920	1133	1123	-9.7	25,900	1010	1386	745	-6.4	41,200
832	1073	1039	-10.7	29,000	921	1123	380	-9.8	71,500	1011	459	541	-29.4	53,500
833	481	820	-28.5	37,800	923	829	242	-15.6	113,200	1012	679	661	-19.7	45,600
834	501	581	-27.8	50,500	924	1131	318	-9.7	84,300	1013	1818	1128	-0.9	25,800
837	751	748	-17.6	41,100	925	1441	874	-5.5	35,400	1014	1032	634	-11.4	47,200
838	635	833	-21.3	37,200	926	679	219	-19.7	128,200	1015	1629	994	-3.0	30,700
839	1494	459	-4.7	60,900	927	1487	1191	-4.8	23,500	1016	1311	1134	-7.4	25,500
840	1952	301	>0.0	89,300	928	1082	775	-10.5	39,800	1017	1722	424	-2.0	65,000
841	1585	1080	-3.6	27,500	929	1231	816	-8.4	38,000	1018	1015	743	-11.7	41,300
842	571	1312	-24.1	19,400	931	1609	670	-3.3	45,100	1020	1574	1219	-3.7	22,500
843	1325	649	-7.2	46,300	932	810	900	-16.0	34,400	1021	781	484	-16.8	58,400
844	1727	301	-2.0	89,200	933	965	520	-12.8	55,100	1022	1129	83	-9.7	50,300
845	630	679	-21.5	44,600	934	947	462	-13.2	60,600	1023	812	317	-15.9	84,600
846	2016	905	>0.0	34,200	936	865	843	-14.8	36,800	1024	785	446	-16.7	62,400
847	673	1200	-19.9	23,200	937	1421	1056	-5.9	28,400	1025	1290	739	-7.7	41,500

MSN	X	Y	CPKd	SDSMW	MSN	X	Y	CPKd	SDSMW	MSN	X	Y	CPKd	SDSMW
1226	405	552	-32.5	52,600	1153	921	1158	-13.7	24,700	1246	547	577	-25.3	50,800
1227	1298	848	-7.5	36,500	1154	1594	864	-3.5	35,900	1247	530	576	-26.3	50,800
1228	856	547	-15.0	53,000	1161	637	400	-21.3	68,400	1248	516	572	-27.0	51,200
1229	1284	226	-7.7	123,200	1162	623	397	-21.8	68,800	1250	973	536	-12.7	53,800
1231	986	822	-12.3	37,700	1163	665	397	-20.2	68,700	1251	607	532	-22.4	54,200
1232	1547	403	-4.1	67,900	1168	564	528	-24.4	54,500	1252	665	529	-20.2	54,400
1233	1381	551	-6.4	52,700	1170	552	529	-25.0	54,500	1253	899	766	-14.1	40,200
1234	1525	496	-4.3	57,200	1171	538	524	-25.9	54,800	1254	1311	746	-7.4	41,200
1235	1128	645	-9.7	46,500	1172	545	514	-25.5	55,700	1255	1300	761	-7.5	40,400
1236	1226	274	-8.5	98,300	1174	1099	522	-10.2	55,000	1257	1938	712	0.0	42,900
1239	1761	262	-1.6	103,600	1176	1304	586	-7.5	50,200	1258	1806	718	-1.0	42,600
1040	541	839	-25.7	36,900	1177	1366	539	-6.6	53,700	1259	1727	715	-2.0	42,700
1041	818	910	-15.8	34,000	1178	1606	702	-3.3	43,400	1260	1629	713	-3.0	42,800
1044	1036	485	-11.3	58,300	1179	1485	224	-4.8	124,900	1261	1555	717	-4.0	42,600
1045	1439	407	-5.5	67,300	1180	1459	224	-5.2	124,900	1262	1468	717	-5.0	42,600
1047	1540	250	-4.2	108,200	1181	1431	223	-5.7	125,100	1263	1413	722	-6.0	42,400
1048	1576	635	-3.7	47,100	1182	1407	223	-6.1	125,200	1264	1340	717	-7.0	42,600
1049	1089	411	-10.4	66,700	1183	1383	224	-6.4	124,700	1265	1263	717	-8.0	42,600
1050	949	1040	-13.2	28,900	1184	1454	182	-5.3	164,400	1266	1182	720	-9.0	42,500
1051	426	818	-31.1	37,800	1185	1422	183	-5.8	162,600	1267	1110	717	-10.0	42,600
1052	1583	1385	-3.6	16,900	1186	1394	182	-6.3	164,300	1268	1055	717	-11.0	42,600
1053	779	1092	-16.8	27,000	1189	1171	214	-9.2	131,800	1269	999	717	-12.0	42,600
1054	1613	620	-3.2	48,000	1190	1457	296	-5.2	94,200	1270	959	715	-13.0	42,700
1055	1380	377	-6.5	72,000	1191	686	1114	-19.5	26,200	1271	805	712	-14.0	42,900
1056	284	663	<-35.0	45,500	1192	265	893	<-35.0	34,700	1272	857	714	-15.0	42,800
1058	1261	746	-8.0	41,200	1193	403	1292	-32.6	20,000	1273	810	705	-16.0	43,300
1060	393	605	-33.3	49,000	1194	344	1275	<-35.0	20,600	1274	774	711	-17.0	42,900
1061	1817	645	-0.9	46,600	1195	505	1311	-27.6	19,400	1277	737	708	-18.0	43,100
1062	1245	746	-8.2	41,200	1196	572	1293	-24.1	20,000	1278	702	711	-19.0	42,900
1064	1258	792	-8.1	39,000	1197	639	1502	-21.2	13,000	1279	671	710	-20.0	43,000
1065	705	934	-18.9	33,000	1198	637	1402	-21.3	16,300	1280	645	710	-21.0	43,000
1066	1181	734	-9.0	41,800	1199	614	1407	-22.1	16,200	1281	617	707	-22.0	43,100
1067	529	658	-26.3	45,800	1200	637	1431	-21.3	15,400	1282	595	704	-23.0	43,300
1068	508	696	-27.4	43,700	1201	1095	1394	-10.3	16,600	1283	573	700	-24.0	43,500
1069	1898	604	-0.3	49,100	1202	1719	1545	-2.1	11,600	1284	552	695	-25.0	43,700
1071	873	609	-14.7	48,700	1203	791	668	-16.5	45,200	1285	536	694	-26.0	43,800
1073	1768	1128	-1.5	25,800	1204	964	1021	-12.9	29,700	1286	515	687	-27.0	44,200
1075	836	773	-15.4	39,900	1205	313	195	<-35.0	148,700	1287	496	683	-28.0	44,400
1076	1863	861	-0.6	36,000	1206	306	194	<-35.0	149,800	1288	467	669	-29.0	45,200
1078	826	566	-15.7	51,600	1209	320	197	<-35.0	147,400	1289	447	667	-30.9	45,300
1081	971	483	-12.7	58,500	1210	326	197	<-35.0	146,600	1290	427	655	-31.0	45,900
1083	1697	202	-2.3	142,300	1211	394	294	-33.2	91,400	1291	412	655	-32.0	45,900
1085	1157	794	-9.4	38,900	1212	402	294	-32.7	91,200	1292	397	652	-33.0	46,100
1090	620	910	-21.9	34,000	1214	386	294	-33.7	91,400	1293	381	654	-34.0	46,000
1092	1867	597	-0.5	49,500	1215	641	329	-21.2	81,600	1294	365	653	-35.0	46,100
1093	2019	894	>0.0	34,600	1216	660	329	-20.4	81,600	1295	348	653	<-35.0	46,100
1094	1546	538	-4.1	53,700	1217	914	266	-13.8	101,800					
1095	1545	477	-4.1	59,100	1218	873	245	-14.7	112,000					
1096	61	935	<-35.0	33,000	1219	970	372	-12.7	72,900					
1099	1954	237	>0.0	116,000	1220	1021	298	-11.6	90,100					
1101	588	1048	-23.3	28,600	1221	1392	205	-6.3	139,500					
1102	1050	657	-11.1	45,200	1222	1354	203	-6.8	141,800					
1103	457	797	-29.5	38,800	1223	1362	205	-6.7	139,500					
1105	1884	532	-0.4	54,200	1224	573	540	-19.9	53,600					
1106	1714	649	-2.1	46,300	1225	614	542	-22.1	53,400					
1107	1717	546	-2.1	53,100	1226	603	539	-22.6	53,600					
1108	1976	722	>0.0	42,400	1227	696	623	-19.2	47,800					
1111	547	1066	-25.3	28,000	1228	707	628	-18.9	47,500					
1112	1348	621	-6.9	48,000	1229	475	447	-28.7	62,300					
1115	1385	762	-6.4	40,400	1230	466	1282	-29.0	20,400					
1116	1078	816	-10.6	38,000	1231	759	1461	-17.4	14,400					
1117	975	787	-12.6	39,300	1232	1324	1170	-7.2	24,200					
1118	1202	933	-8.7	33,100	1233	1583	1005	-3.6	30,300					
1119	1022	1076	-11.6	27,600	1234	1865	809	-0.6	38,200					
1120	1905	616	-0.3	48,300	1235	1812	817	-1.0	37,900					
1121	1512	1301	-4.5	19,700	1236	1411	703	-6.0	43,400					
1122	1114	677	-9.9	44,700	1237	1392	682	-6.3	44,500					
1123	1464	452	-5.1	61,700	1238	794	410	-16.4	66,900					
1125	1048	857	-11.1	36,200	1239	769	407	-17.1	67,300					
1126	1122	802	-9.8	38,600	1240	740	406	-17.9	67,500					
1128	1722	892	-2.1	34,700	1241	743	511	-17.8	55,900					
1133	1098	825	-10.2	37,500	1242	713	510	-18.7	56,000					
1139	1630	569	-0.8	51,400	1243	682	509	-19.6	56,100					
1147	764	1182	-17.3	23,800	1244	663	504	-20.3	56,500					
1148	1968	724	>0.0	42,300	1245	565	582	-24.4	50,500					

Table 2. Table of some identified proteins
POI^a name PEST content^b

POI ^a name	PEST content ^b	MSN's	Basis for identification
IDS:3_ALPHA_HDHD	3- α -hydroxysteroid-dihydrodol-dehydrogenase, an enzyme of steroid metabolism	197, 159	Pure protein and antibody provided by Dr. T.M. Penning, Department of Pharmacology, School of Medicine, University of Pennsylvania.
IDS:ACTIN_BETA	β cellular actin, a cytoskeletal protein	38	Homologous position with respect to other mammalian systems
IDS:ACTIN_GAMMA	γ cellular actin, a cytoskeletal protein	68	Homologous position with respect to other mammalian systems
IDS:ALBUMIN	Serum albumin, mature form.	21, 28, 33	Predominance in rat plasma
IDS:APO_A-I	Apo A-I plasma lipoprotein, mature form (fetaline)	238, 463	Presence in rat plasma, regulation by some lipid-lowering drugs
IDS:CALMODULIN	Calmodulin, an acidic cytosolic calcium-binding protein	123, 649	Homologous position with respect to other mammalian systems
IDS:CATALASE	Catalase (peroxisomal)	54, 61, 106	Presence in purified peroxisomes, similarity in position to mouse catalase
IDS:CPKSPOTS	Spots contributed by the CPK charge standards (not rat liver proteins)	1257 - 1295	
IDS:CPS	Carbamoyl phosphate synthase	87, 477	
IDS:CYTOCHROME_B5	Cytochrome b5		114, 157, 167, 174, 1164, 1185, 1166, 1222 Pure protein provided by Dr. Margaret Marshall, Department of Pharmacology, Medical School, University of Wisconsin - Madison.
IDS:FABP_L	Liver fatty-acid binding protein	.227	Pure protein provided by Dr. Naihan Bass, Department of Medicine, University of California School of Medicine, San Francisco
IDS:HMG_COA_SYNTHASE	Cytosolic HMG-CoA Synthase	133, 144, 235, 413	Antibody provided by Dr. Michael Greenspan, Merck Sharp & Dome Research Laboratories, Rahway, NJ
IDS:LAMIN_B	Lamin B, a nuclear protein	415, 734	Homologous position with respect to other mammalian systems
IDS:MITCON_1	Mitcon:1 (F1 ATPase β subunit), a mitochondrial inner membrane protein	17, 49, 71, 340, 1245, 1246, 1247, 1249	Homologous position with respect to other mammalian systems, presence in mitochondria
IDS:MITCON_2	Mitcon:2, a mitochondrial matrix stress protein equivalent to E.	15, 25, 110, 1241, 1242, 1243, 1244	Homologous position with respect to other mammalian systems, presence in mitochondria
IDS:MITCON_3	Mitcon:3, a mitochondrial matrix stress protein, likely analog of NADPH cyochrome P-450 reductase	18, 35, 226, 600, 1238, 1239, 1240	Homologous position with respect to other mammalian systems, presence in mitochondria
IDS:NADPH_P450_RED	NADPH cyochrome P-450 reductase, frequently co-induced with P-450's	175, 251, 812	Pure protein provided by Dr. Andrew Parkinson, Department of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center
IDS:PDI	Protein disulphide isomerase 1	168, 1170, 1171, 1172	Sequence information obtained by R.M. Van Frank, Lilly Research Laboratories, Indianapolis
IDS:PLASMA_PROTEINS	Rat plasma proteins observed in liver	21, 28, 33, 44, 72, 102, 115, 197, 236, 246, 738, 790, 865, 903, 926, 47, 93	Relative position to mature albumin, presence in micro-somes
IDS:PRO_ALBUMIN	Serum albumin precursor	179, 1180, 1181, 1182, 1183	Pavlica, R.J., et al., BBA (1990) 1022 115-125.
IDS:PYRCARBOX	Pyruvate carboxylase	135	Sequence information obtained by R.M. Van Frank, Lilly Research Laboratories, Indianapolis
IDS:SOD	Superoxide dismutase		Homologous position with respect to other mammalian systems
IDS:TUBULIN_ALPHA	α tubulin, a cytoskeletal protein	56, 132, 1224, 1232	Homologous position with respect to other mammalian systems
IDS:TUBULIN_BETA	β tubulin, a cytoskeletal protein	50, 1225, 1226, 1251	

Hb-beta

Computed hemoglobin
Protein
Rabbit r

Table 3. Computed pIs of two sets of carbamylated protein standards: Rabbit muscle CPK and human hemoglobin (Hb)

Protein Name	PIR Name	#ASP							Calc pI	Real CPK
		3.9	4.1	6.0	10.8	12.5	NH2- 7.0			
Rabbit muscle CPK	KIRBCM	28	27	17	34	18	1	6.84	0.0	
		28	27	17	33	18	1	6.67	-1	
		28	27	17	32	18	1	6.54	-2	
		28	27	17	31	18	1	6.42	-3	
		28	27	17	30	18	1	6.31	-4	
		28	27	17	29	18	1	6.21	-5	
		28	27	17	28	18	1	6.12	-6	
		28	27	17	27	18	1	6.03	-7	
		28	27	17	26	18	1	5.94	-8	
		28	27	17	25	18	1	5.85	-9	
		26	27	17	24	18	1	5.76	-10	
		28	27	17	23	18	1	5.67	-11	
		28	27	17	22	18	1	5.58	-12	
		28	27	17	21	18	1	5.48	-13	
		28	27	17	20	18	1	5.39	-14	
		28	27	17	19	18	1	5.29	-15	
		28	27	17	18	18	1	5.20	-16	
		28	27	17	17	18	1	5.12	-17	
		28	27	17	16	18	1	5.04	-18	
		28	27	17	15	18	1	4.96	-19	
		28	27	17	14	18	1	4.89	-20	
		28	27	17	13	18	1	4.83	-21	
		28	27	17	12	18	1	4.77	-22	
		28	27	17	11	18	1	4.71	-23	
		28	27	17	10	18	1	4.66	-24	
		28	27	17	9	18	1	4.61	-25	
		28	27	17	8	18	1	4.56	-26	
		28	27	17	7	18	1	4.52	-27	
		28	27	17	6	18	1	4.48	-28	
		28	27	17	5	18	1	4.44	-29	
		28	27	17	4	18	1	4.40	-30	
		28	27	17	3	18	1	4.36	-31	
		28	27	17	2	18	1	4.32	-32	
		28	27	17	1	18	1	4.29	-33	
		28	27	17	0	18	1	4.25	-34	
		28	27	17	0	18	0	4.22	-35	
Hb-beta, human	HBHU	7	8	9	11	3	1	7.18		
		7	8	9	10	3	1	6.79		
		7	8	9	9	3	1	6.53	-1.8	
		7	8	9	8	3	1	6.32	-3.2	
		7	8	9	7	3	1	6.13	-5.3	
		7	8	9	6	3	1	5.96	-7.2	
		7	8	9	5	3	1	5.78	-10.0	
		7	8	9	4	3	1	5.59	-12.3	
		7	8	9	3	3	1	5.37	-15.5	
		7	8	9	2	3	1	5.14	-18.0	
		7	8	9	1	3	1	4.91	-21.0	
		7	8	9	0	3	1	4.71	-25.5	
		7	8	9	0	3	0	4.54	-27.2	

Table 4. Computed pIs of some known proteins related to measured CPK pIs

Protein Name	PIR Name	#ASP #GLU #HIS #LYS #ARG						Calc pI	Real CPK
		3.9	4.1	6.0	10.8	12.5			
Creatine phospho kinase (CPK), rabbit muscle	KIRBCM	28	27	17	34	18	6.84	0.0	
Fatty acid-binding protein, rat hepatic	FZRTL	5	13	2	16	2	7.83	-3.0	
b2-microglobulin, human	MGHUB2	7	8	4	8	5	6.09	-5.0	
Carbamoyl-phosphate synthase, rat	SYRTCA	72	96	28	95	56	5.97	-5.5	
Proalbumin (serum albumin precursor), rat	ABRTS	32	57	15	53	27	5.98	-6.2	
Serum albumin, rat	ABRTS	32	57	15	53	24	5.71	-9.0	
Superoxid dismutase (Cu-Zn. SOD), rat	A26810	8	11	10	9	4	5.91	-9.2	
Phospholipase C, phosphoinositide-specific (?), rat	A28807	34	42	9	49	21	5.92	-9.2	
Albumin, human	ABHUS	36	61	16	60	24	5.70	-11.9	
Apo A-I lipoprotein, rat	A24700	18	24	6	23	12	5.32	-13.7	
proApo A-I lipoprotein, human	LPHUA1	16	30	6	21	17	5.35	-14.3	
NADPH cytochrome P-450 reductase, rat	RDRTO4	41	60	21	38	36	5.07	-15.6	
Retinol binding protein, human	VAHU	18	10	2	10	14	5.04	-16.9	
Actin beta, rat	ATRTC	23	26	9	19	18	5.06	-17.2	
Actin gamma, rat	ATRTC	20	29	9	19	18	5.07	-16.8	
Apo A-I lipoprotein, human	LPHUA1	16	30	5	21	16	5.10	-17.5	
Apo A-IV lipoprotein, human	LPHUA4	20	49	8	28	24	4.88	-19.7	
Tubulin alpha, rat	UBRTA	27	37	13	19	21	4.66	-19.8	
F1ATPase beta, bovine	PWBGB	25	36	9	22	22	4.80	-21.0	
Tubulin beta, pig	UBPGB	26	36	10	15	22	4.49	-22.5	
Protein disulphide isomerase (PDI), rat hepatic	ISRTSS	43	51	11	51	9	4.07	-25.0	
Cytochrome b5, rat	CBRT5	10	15	6	10	4	4.59	-26.0	
Apo C-II lipoprotein, human	LPHUC2	4	7	0	6	1	4.44	-30.5	
Amino acid pI assumed in calculation:		3.9	4.1	6.0	10.8	12.5			

Wirth
Loo
Fujimoto
C. Bisgaard
D. Olson

Institute of Exper.
ogenesis,
National Cancer Inst.
Institutes
Canada,

Introduction
Materials and
Materials
Cells.....
Metabolic
nine and
Sample p:
Subcellul:
2-D PAG:
Computer
retograms
Results
[³⁵S]Methi
Whole ce
Subcellul:
[³²P]Ortho
Discussion...
References...
Appendix 1:
Appendix 2:
Proteins

Correspondence: Dr. P
National Car
32, USA

Address: 2-D PA
HLE, hum:
ar weight; NE
Nonidet P
ode; RLE, rat

Engagegesellschaft:

N. Leigh Anderson¹
 Ricardo Esquer-Blasco¹
 Jean-Paul Hofmann¹
 Lydie Meheus²
 Jos Raymackers²
 Sandra Steiner³
 Frank Wizmann³
 Norman G. Anderson¹

¹Large Scale Biology Corporation,
 Rockville, MD
²Innogenetics NV, Ghent
³Sandoz Pharma Ltd, Drug Safety
 Assessment, Toxicology, Basel
⁴Molecular Anatomy Laboratory,
 Indiana University Purdue
 University Columbus, Columbus, IN

An updated two-dimensional gel database of rat liver proteins useful in gene regulation and drug effect studies

We have improved upon the reference two-dimensional (2-D) electrophoretic map of rat liver proteins originally published in 1991 (N. L. Anderson *et al.*, *Electrophoresis* 1991, 12, 907-930). A total of 53 proteins (102 spots) are now identified, many by microsequencing. In most cases, spots cut from wet, Coomassie Blue stained 2-D gels were submitted to internal tryptic digestion [2], and individual peptides, separated by high-performance liquid chromatography (HPLC), were sequenced using a Perkin-Elmer 477A sequenator. Additional spots were identified using specific antibodies.

Figure 1 shows the current annotated 2-D map of F344 rat liver, analyzed using the Iso-DALT system (20 × 25 cm gels) and BDH 4-8 carrier ampholytes. Both the map itself and the master spot number system remain the same as shown in the original publication. Table 1 lists the important features of each identification shown, including the gel position, *pI*, and *M*, for the most abundant or most basic form of each protein. Using this extended base of identified spots, a series of four improved calibration functions has been derived for the *pI* and SDS-*M*, axes (the first two of which are shown in Fig. 2A and B). Both forward and reverse functions are derived, so that one can compute the physical properties of a spot with a given gel location, or inversely compute the gel position expected for a protein having given physical properties:

$$Y_{\text{RATLIVER}} = f_{M-\text{RATLIVER}} \circ (M_{\text{SEQUENCE-DERIVED}}) \quad (1)$$

$$X_{\text{RATLIVER}} = f_{pI-\text{RATLIVER}} \circ (pI_{\text{SEQUENCE-DERIVED}}) \quad (2)$$

$$M_{\text{GEL-DERIVED}} = f_{\text{RATLIVER } Y-M} \circ (Y_{\text{RATLIVER}}) \quad (3)$$

$$pI_{\text{GEL-DERIVED}} = f_{\text{RATLIVER } X-pI} \circ (X_{\text{RATLIVER}}) \quad (4)$$

A spreadsheet program (in Microsoft Excel) was developed to facilitate flexible computation of *pI*'s from amino acid sequence data, and the results were entered into a relational database (Microsoft Access). A table of spot positions and sequence-derived *pI*'s and *M*'s was fitted with a large series of analytic equations using Tablecurve (Jandel Scientific), and the four conversion Eqs. (1)-(4), relating computed *pI* and gel *X* coordinate, or computed molecular weight and gel *Y* coordinate, were selected, based on criteria of simplicity, goodness of fit and favorable asymptotic behavior. Table 2 lists the equations and coefficients. Application of Eqs. (3) and (4) to a spot's *X* and *Y* coordinates, given in (1), produce improved *M*, estimates, and allow computation of *pI*/

directly in pH units, instead of in terms of positions relative to creatine phosphokinase (CPK) charge standards. The inverse Eqs. (1) and (2) were used to compute the gel positions of a series of *pI* and *M*, tick marks. These tick marks were plotted with SigmaPlot (Jandel), together with fiducial marks locating several prominent spots, and the resulting graphic was aligned over the synthetic gel image (computed by Kepler from the master gel pattern) using Freelance (Lotus Development). Maps were printed as Postscript output from Freelance, either in black and white (as shown here) or in color, where label color indicates subcellular location (available from the first author upon request). We have also used the rat liver 2-D pattern as presented here to calibrate the patterns of other samples. Using mixtures of rat liver and mouse liver samples, for example, we made composite 2-D patterns that allow use of the rat pattern to standardize both axes of the mouse pattern. This was accomplished by deriving transformations relating the rat and mouse *X*, and separately the rat and mouse *Y*, axes (Table 2, lower half; Fig. 2C and D) based on a series of spots that coelectrophorese in these closely related species. These functions were then applied to derive equations relating the mouse liver *X* and *Y* to *pI* and SDS-*M*, (Eqs. 5 and 6 below). The resulting standardized 2-D pattern for B6C3F1 mouse liver is shown in Fig. 3.

$$M_{\text{MOUSELIVER}} = f_{\text{RATLIVER } Y-M} \circ (f_{\text{MOUSELIVER } Y-\text{RATLIVER } Y} \circ (Y_{\text{MOUSELIVER}})) \quad (5)$$

$$pI_{\text{MOUSELIVER}} = f_{\text{RATLIVER } X-pI} \circ (f_{\text{MOUSELIVER } X-\text{RATLIVER } X} \circ (X_{\text{MOUSELIVER}})) \quad (6)$$

A slightly more complex approach can be used to standardize samples that have few or no spots co-electrophoresing with rat liver proteins. In this case, a 2-D gel is prepared with a mixture of the two samples, and four functions (forward and backward, each for *X* and *Y*) are derived relating each sample's own master pattern to the composite. The required functions are then applied in a nested fashion to yield the desired result (using rat plasma as an example):

$$M_{\text{RATPLASMA}} = f_{\text{RATLIVER } Y-M} \circ (f_{\text{RATPLASMA+LIVER } Y-\text{RATLIVER } Y} \circ (f_{\text{RATPLASMA } Y-\text{RATPLASMA+LIVER } Y} \circ (Y_{\text{RATPLASMA}}))) \quad (7)$$

SANTANDER LIBRARIES

Correspondence: Dr. Leigh Anderson, Large Scale Biology Corporation, 9620 Medical Center Drive, Rockville, MD 20850-3338 USA (Tel: +301-424-5989; Fax: +301-762-4892; email: leigh lsbc.com)

Keywords: Two-dimensional polyacrylamide gel electrophoresis / Liver / Map / Identification / Calibration

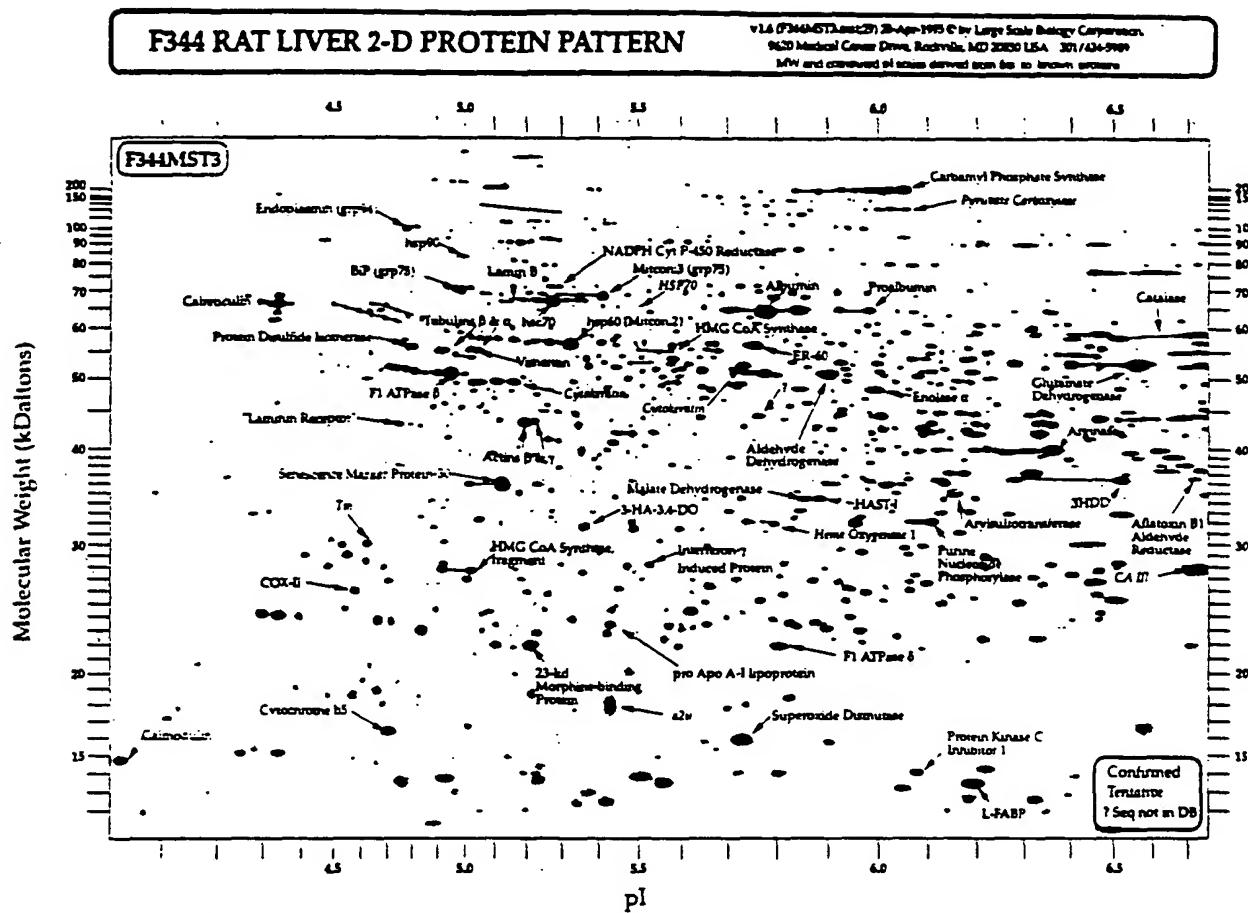


Figure 1. Master 2-D gel pattern of Fischer 344 rat liver proteins, annotated with 53 protein identifications and computed pI and M_r axes. Tentative identifications are in italic type.

Table 1. Proteins identified in the 2-D pattern of F344 rat liver

MSN ^{a)}	Protein ID ^{b)}	Protein name	Identification comments	Gel X ^{c)}	Experimental pI ^{d)}	Gel Y ^{c)}	Experimental M _r ^{d)}
126	HADO-HUMAN*	3-HA-3,4-DO: 3-hydroxy-anthraniolate-3,4-dioxygenase	Internal sequence	871.95	5.36	921.35	30 207
137, 159, 288, 258	DIDH_RAT	3HDD: 3-hydroxysteroid dihydrodiol reductase	Ab (T.M. Penning) and pure protein	1857.52	6.51	822.52	34 406
173	MUP_RAT	α_2u globulin	Presence in liver microsome lumen, abundance in kidney, pI, M _r	919.16	5.43	1313.81	19 549
38	ACTB_HUMAN	Actin β	Analogy with other mammalian patterns (e.g. human) through coelectrophoresis	763.40	5.19	693.64	41 586
68	ACTG_HUMAN	Actin γ	Analogy with other mammalian patterns (e.g. human) through coelectrophoresis	779.42	5.21	692.26	41 677
693	AFAR_RAT	Aflatoxin B1 aldehyde reductase	Internal sequence	1993.32	6.72	818.60	34 593
28, 21, 33	ALBU_RAT	Albumin	Coelectrophoresis with principal plasma protein	1262.81	5.86	445.64	66 354
43	DHAM_RAT	Aldehyde dehydrogenase	N-Terminal sequence and AAA	1317.72	5.91	589.03	49 602
96	ARG1_RAT	Arginase	Internal sequence	1730.72	6.34	756.02	37 819
117	SUAR_RAT	Arylsulfotransferase	Internal sequence	1547.96	6.14	849.08	33 186
1163, 1161, 1162, 20	GR78_RAT	BIP (GRP-78)	Ab (F. Wilzmann)	665.33	5.01	397.39	74 564
185	CAH3_RAT	CA-III	Uncertain; by comparison with mouse	1996.60	6.72	1017.02	26 887
123	CALM_HUMAN	Calmodulin	Analogy with human cellular patterns through coelectrophoresis	23.05	4.03	1433.25	17 419
3, 201, 48, 39, 22, 24	CRTC_RAT	Calreticulin	Ab (Lance Pohl)	310.59	4.34	433.80	68 206

Table 1. continued

MSN ^{a)}	Protein IDb)	Protein name	Identification comments	Gel X ^{c)} pI ^{d)}	Experimental pI ^{d)}	Gel Y ^{c)} M _r ^{e)}	Experimental M _r ^{e)}
1184, 1186, 114, 174, 118 5, 167, 157	CPSM_RAT	Carbamyl phosphate synthase	2-D of pure protein; confirmed by <i>N</i> -terminal sequence and AAA	1453.56	6.05	181.64	160 640
54, 61 136	CATA_RAT COX2_RAT	Catalase COX-II	Internal sequence Ab (J. W. Taanman), confirmed by internal sequence	2000.81 452.57	6.73 4.61	499.64 1062.67	58 968 25 504
87	CYBS_RAT	Cytochrome BS	2-D of pure protein; Ab; confirmed by AAA	515.68	4.73	1370.55	18 493
41 29 5, 11 60 27 17 196 79	CK-RAT ^{d)} CK-RAT ^{d)} ENPL-RAT ^{d)} ENO_A_RAT ER60_RAT ATPB_RAT ATP7_RAT F16P_RAT	Cytokeratin Cytokeratin Endoplasmic Enolase A ER-60 F1 ATPase β F1 ATPase δ Fructose-1,6-bis-phosphatase	Location in cytoskeletal fraction Location in cytoskeletal fraction Ab (F. Witzmann) Internal sequence and AAA <i>N</i> -Terminal sequence (R. M. Van Frank) <i>N</i> -Terminal sequence and AAA Internal sequence Uncertain; by comparison with ID in Garrison and Wager (JBC 257:13135-13143)	1165.12 743.11 567.73 1399.78 1184.20 629.06 1227.24 924.54	5.75 5.15 4.83 6.00 5.77 4.95 5.82 5.44	569.09 605.23 263.37 623.54 523.51 588.83 1184.65 737.77	51 448 48 187 112 194 46 674 56.169 49 620 22 310 38 858
62, 78 125	DHE3_RAT HAST-RAT ^{d)}	Glutamate dehydrogenase HAST-1: N-hydroxyarylamine sulfotransferase	<i>N</i> -Terminal sequence and internal sequence Internal sequence	1887.39 1297.94	6.55 5.89	566.92 861.55	51 655 32 638
307	HO1_RAT	Heme oxygenase 1	Uncertain; available data from internal sequence	1219.39	5.81	915.71	30 423
413, 1250, 933 133, 144, 235	HMCS_RAT HMCS_RAT	HMG CoA synthase, cytosolic HMG CoA synthase, mitochondrial (frag)	Ab (J. Germershausen) Ab (J. Germershausen), <i>N</i> -terminal sequence (Steiner/Lotuspeich)	1033.48 666.40	5.59 5.02	538.13 1019.42	54 571 26 811
8, 23, 1307	HS7C_RAT	HSC-70	Positional homology (with human, etc.) through coelectrophoresis	811.87	5.27	425.76	69 521
15, 25, 110	P60_RAT	HSP-60	Ab (F. Witzmann); confirmed by <i>N</i> -terminal sequence and AAA	845.09	5.32	520.03	56 561
971 1216, 1215, 90 256	HS70-RAT ^{d)} HS90-RAT ^{d)} INGI-HUMAN	HSP-70 HSP-90 Interferon- γ induced protein	Ab (F. Witzmann) Ab (F. Witzmann) Internal sequence	976.11 659.86 993.85	5.51 5.00 5.54	437.14 329 1006.04	67 674 90 107 27 237
415, 734	LAMB-RAT ^{d)}	Lamin B	Positional homology with human through coelectrophoresis, nuclear location	737.10	5.14	425.19	69 615
80 227	LAMR-RAT ^{d)} FABL_RAT	"Laminin receptor" L-FABP (liver fatty acid binding protein)	Internal sequence Ab (N. M. Bass)	534.02 1586.09	4.77 6.18	697.62 1483.43	41 327 16 622
134	MDHC_MOUSE	Malate dehydrogenase	Internal sequence	1270.85	5.86	861.96	32 620
18, 35, 226	GR75-RAT ^{d)}	Mitcon-3; grp75	Positional homology with human through coelectrophoresis	905.67	5.41	413.67	71 589
175, 251 1168, 1170, 1171	NCPR_RAT PDI_RAT	NADPH P450 reductase PDI: Protein disulfide isomerase	2-D of pure protein <i>N</i> -Terminal sequence (R. M. van Frank), Ab	824.69 564.30	5.29 4.83	393.21 528.47	75 366 55 618
47, 93	ALBU_RAT	Pro-Albumin	Microsomal lumen location, pI, M _r , relative to albumin	1391.03	5.99	446.68	66 195
236 320	APA1_RAT IPK1_BOVIN	Pro-APO A-I lipoprotein Protein kinase C inhibitor 1	Colectrophoresis with plasma protein Internal sequence; homology with bovine protein	920.41 1480.01	5.43 6.08	1137.51 1458.81	23 467 17 007
152	PNPH_MOUSE	Purine nucleoside phosphorylase	Internal sequence	1507.19	6.10	911.16	30 599
1179, 1180, 1181, 1182, 1183 55, 103	PYVC-RAT ^{d)} SM30_RAT	Pyruvate carboxylase SMP-30: Senescence marker protein-30	Tentative; 2-D of pure protein (J. G. Henslee, JBC, 1979); reported in <i>Biochim. Biophys. Acta</i> 1022, 115-125.	1485.10	6.08	223.52	131 589
135	SODC_RAT	Superoxide dismutase	AAA; confirmed by internal sequence (R. M. Van Frank)	1161.24	5.74	1388.68	18 173
172	TPM-RAT ^{d)}	Tm: tropomyosin	Location in cytoskeleton, 2-D position relative to human, Ab	476.24	4.66	957.86	28 865
277, 56	TBA1_RAT	Tubulin α	Positional homology with human through coelectrophoresis, cytoskeletal location	688.22	5.06	537.67	54 620
50, 1225	TBB1_RAT	Tubulin β	Positional homology with human through coelectrophoresis, cytoskeletal location	621.29	4.93	535.48	54 855
1224	VIME_RAT	Vimentin	Positional homology with human through coelectrophoresis, cytoskeletal location	673.00	5.03	539.50	54 426

Table 1. continued

MSN ^{a)}	Protein ID ^{b)}	Protein name	Identification comments	Gel X ^{c)}	Experimental pI ^{d)}	Gel Y ^{c)}	Experimental M _r ^{e)}
813	Unknown	?; not in sequence databases	Internal sequence	1191.28	5.78	680.42	42 469
804	BBPL_RAT	23 kDa morphine-binding protein	Internal sequence	773.31	5.20	1182.41	22 363

^{a)} Master spot number (MSN) from [1]^{b)} SwissPROT identifier^{c)} Coordinates of the most basic or most abundant assigned spot on the F344 master gel pattern^{d)} pI and M_r of the most basic or most abundant assigned spot, derived from the calibration functions included here^{e)} SwissPROT style proposed identifier^{f)} Abbreviations: AAA, amino acid analysis; Ab, antibody

Table 2. Equations and coefficients

action	Equation (f)	r ²	a	b	c	d	e
use gel Y = f(rat M _r) y = a + bexp(-x/c)	0.988181021	178.74803	1967.7892	32363.958			
use gel X = f(rat pI) y = a + bx - cx/lnx - dx/x + ex/x ^{1.5}	0.99247216	-8685665.5	-904497.94	3836926.1	18276844	-27154534	
computed M _r = f(rat gel Y) y = a + bxc	0.9960177	-8464.5809	19095881	-0.9086255			
computed pI = f(rat gel X) y = a + bx + cx ² + dx ^{2.5} lnx + ex ³	0.99176499	4.044686	-0.00114238	0.0000323	-0.00000455	0.00000000176	
use gel Y = f(rat gel Y) y = a + bx + cx ^{1.5} + dx ^{0.5} lnx + ex/lnx	0.99951069	11861.44	678.91666	-0.78964914	1567.5639	-6953.9592	
use gel X = f(rat gel X) y = a + bx ² lnx + cx ^{2.5} + dx ³	0.99926349	58.935923	0.00091353	-0.000213688	0.00000159		
use gel Y = f(mouse gel Y) y = a + bx ² lnx + cx ^{2.5} + dx ³	0.99950032	69.740526	0.00050772	-0.000130392	0.00000116		
use gel X = f(mouse gel X) y = a + bx + cx ² lnx + dx ^{2.5} + ex ³	0.9992832	-198.07189	2.0899063	-0.000671191	0.000145189	-0.000000986	

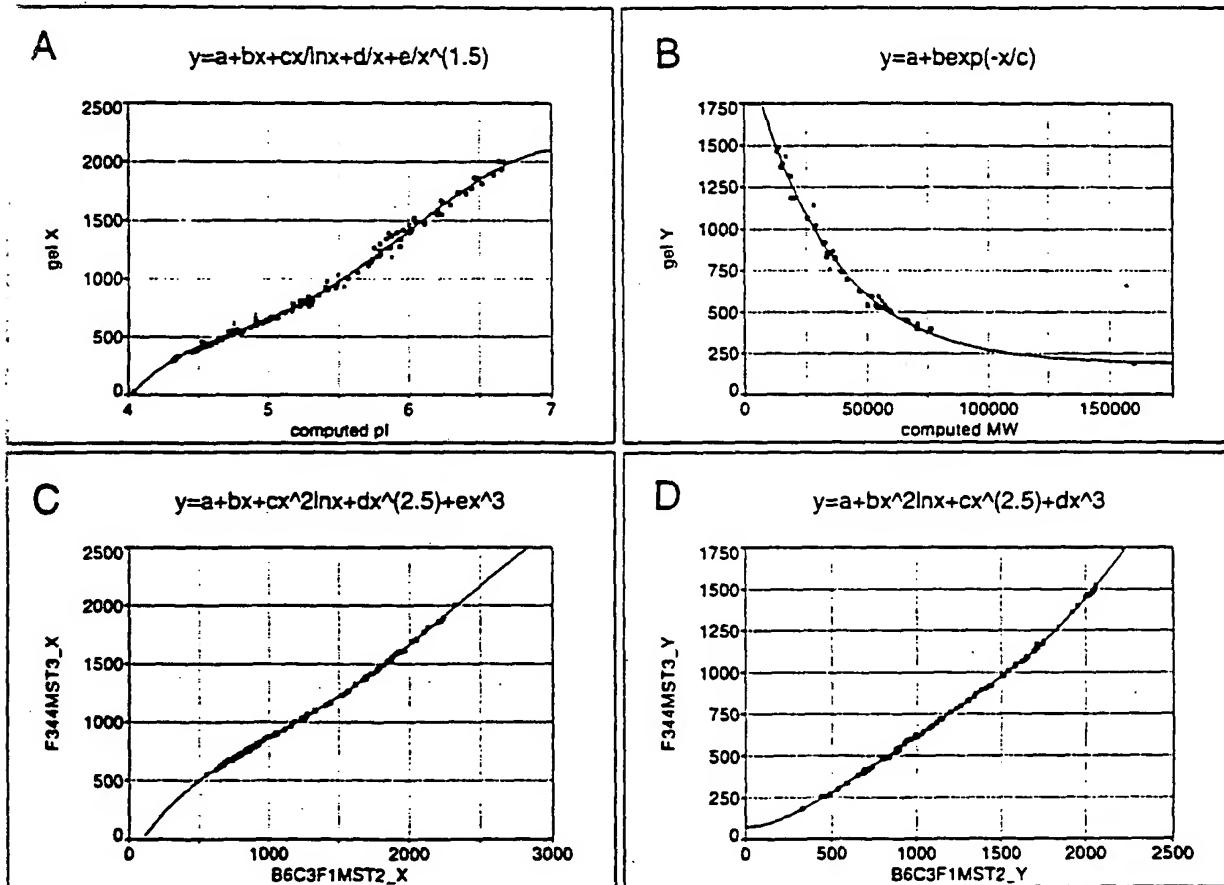


Figure 2. Plots showing fits of selected equations (continuous curves) to data on identified proteins (square symbols). (A) pI computed from sequence data versus gel X position for identified spots in F344 rat liver; (B) M_r computed from sequence data versus gel Y position for identified spots in F344 rat liver; (C) gel X position for spots in B6C3F1 mouse liver versus X position in F344 rat liver, for coelectrophoresing spots; (D) gel Y position for spots in B6C3F1 mouse liver versus Y position in F344 rat liver, for coelectrophoresing spots. In each case, inverse equations were also computed (Table 2).

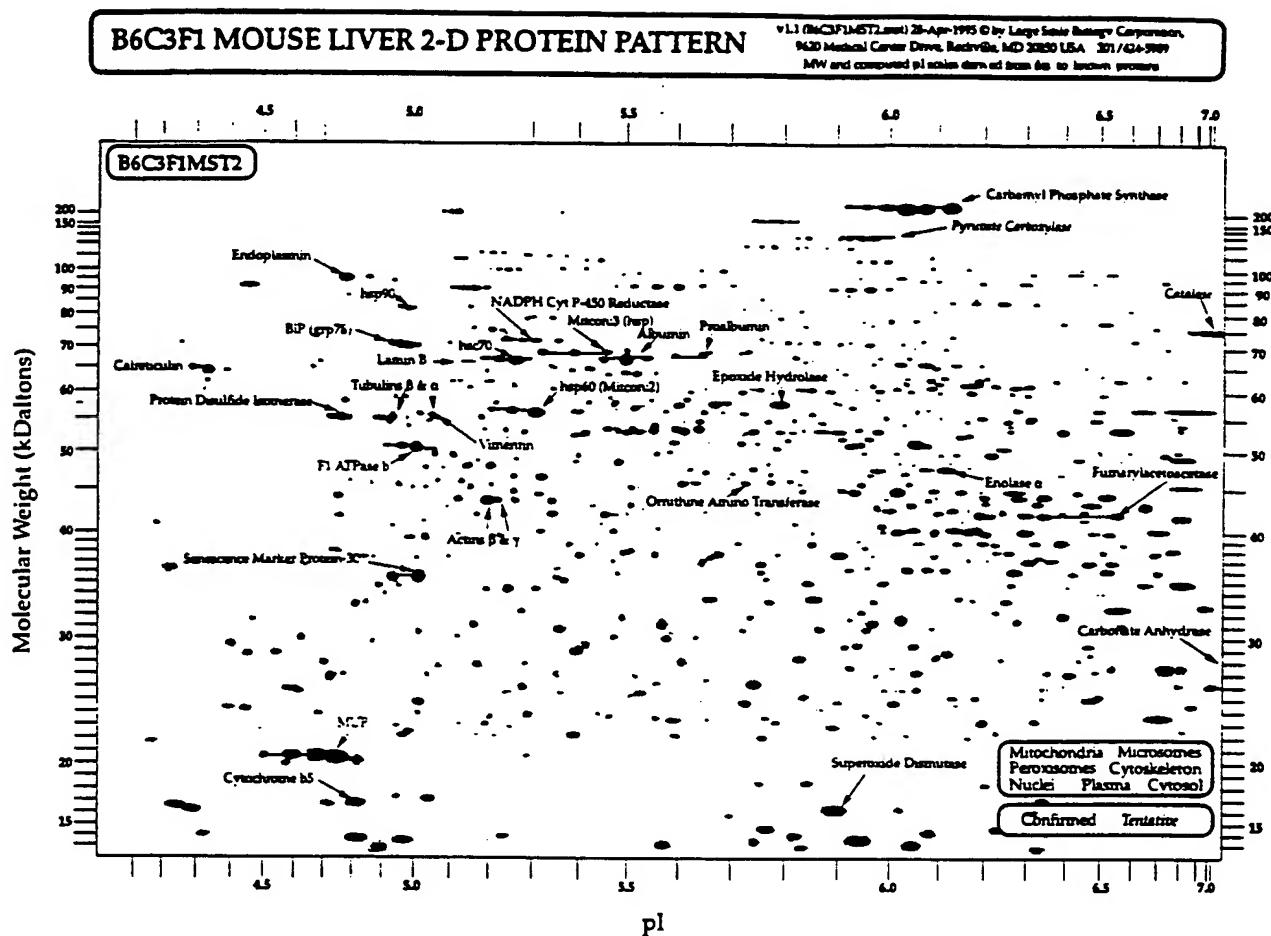


Figure 3. Master 2-D gel pattern for B6C3F1 mouse liver, standardized using the F344 rat liver pattern identifications, according to the method described in the text. Twenty-nine proteins are identified.

$$pI_{RATPLASMA} = f_{RATLIVER} \times pI_{RATPLASMA} + (f_{RATPLASMA} \times LIVER \times (X_{RATLIVER} - X_{RATPLASMA})) / (f_{RATPLASMA} \times RATERPLASMA + LIVER \times (X_{RATPLASMA})) \quad (8)$$

This unified approach, in which one well-populated 2-D pattern is used to standardize a family of other patterns, has the additional advantage that the resulting pI and M scales are directly compatible. Hence one can compare the relative pI 's of mouse and rat versions of a sequenced protein in a consistent pI measurement system, and select likely inter-species analogs based on positional relationships on common scales. Adoption of immobilized pH gradient (IPG) technology [4–7] will result in substantial improvements in pI positional reproducibility for standard 2-D maps such as those presented here; however, we believe that our approach will continue to be useful in establishing the empirical pH gradient actually achieved by such gels under given experimental conditions (temperature, urea concentration, etc.), in relating patterns run on different IPG ranges and using different lots of IPG gels (between which some variation will persist). Development of rodent organ maps is a continuing effort in our laboratories [8–10], and results in regular additions of identified proteins. Those who wish to receive current rodent liver maps, with color annotations, should send a stamped self-addressed envelope to the first author.

We would like to thank the individuals who provided antibodies mentioned in Table I, and R. M. van Frank for unpublished sequenced data.

Received May 31, 1995

References

- [1] Anderson, N. L., Esquer-Blasco, R., Hofmann, J.-P., Anderson, N. G., *Electrophoresis* 1991, 12, 907–930.
- [2] Rosenfeld, J., Capdeville, J., Guillemot, J. C., Ferrara, P., *Anal. Biochem.* 1992, 203, 173–179.
- [3] Witzmann, F., Clack, J., Fultz, C., Jarnot, B., *Electrophoresis* 1995, 16, 451–459.
- [4] Rosengren, A. E., Bjellqvist, B., Gasparic, V., US Patent 4130470, December 1978.
- [5] Gianazza, E., Artoni, G., Righetti, P. G., *Electrophoresis* 1983, 4, 321–326.
- [6] Görg, A., Postel, W., Günther, S., Weser, J., *Electrophoresis* 1985, 6, 599–604.
- [7] Gianazza, E., Astrua-Testori, S., Giaccon, P., Righetti, P. G., *Electrophoresis* 1985, 6, 332–339.
- [8] Myers, T. G., Dietz, E. C., Anderson, N. L., Khairallah, E. A., Cohen, S. D., Nelson, S. D., *Chem. Res. Toxicol.* 1995, 8, 403–413.
- [9] Cunningham, M. L., Pippin, L. L., Anderson, N. L., Wenk, M. L., *Toxicol. Appl. Pharmacol.* 1995, 131, 216–223.
- [10] Anderson, N. L., Coppie, D. C., Bendele, R. A., Probst, G. S., Richardson, F. C., *Fundam. Appl. Toxicol.* 1992, 18, 570–580.



Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It

MARC R. WILKINS¹, JEAN-CHARLES SANCHEZ¹, ANDREW A. GOOLEY¹,
RON D. APPEL¹, IAN HUMPHERY-SMITH², DENIS F. HOCHSTRASSER¹
AND KEITH L. WILLIAMS^{1,*}

¹ Macquarie University Centre for Analytical Biotechnology, Macquarie University, Sydney, NSW 2109, Australia; ² Department of Microbiology, University of Sydney, NSW, 2006, Australia and ¹ Central Clinical Chemistry Laboratory and Medical Computing Centre of the University of Geneva, CH 1211 Geneva 14, Switzerland

Introduction

The advent of large genome sequencing projects has changed the scale of biology. Over a relatively short period of time, we have witnessed the elucidation of the complete nucleotide sequence for bacteriophage λ (Sanger *et al.*, 1982), the nucleotide sequence of an eukaryotic chromosome (Oliver *et al.*, 1992), and in the near future will see the definition of all open reading frames of some simple organisms, including *Mycoplasma pneumoniae*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Arabidopsis thaliana*. Nevertheless, genome sequencing projects are not an end in themselves. In fact, they only represent a starting point to understanding the function of an organism. A great challenge that biologists now face is how the co-expression of thousands of genes can best be examined under physiological and pathophysiological conditions, and how these patterns of expression define an organism.

There are two approaches that can be used to examine gene expression on a large scale. One uses nucleic acid-based technology, the other protein-based technology. The most promising nucleic-acid based technology is differential display of mRNA (Liang and Pardee, 1992; Bauer *et al.*, 1993), which uses polymerase chain reaction with arbitrary primers to generate thousands of cDNA species, each which correspond to an expressed gene or part of a gene. However, it is currently unclear if this technique can be developed to reliably assay the expression of thousands of genes or

* Corresponding Author

identify all cDNA species, and the approach does not easily allow a systematic screening. Analysis of gene expression by the study of proteins present in a cell or tissue presents a favorable alternative. This can be achieved by use of two-dimensional (2-D) gel electrophoresis, quantitative computer image analysis, and protein identification techniques to create 'reference maps' of all detectable proteins. Such reference maps establish patterns of normal and abnormal gene expression in the organism, and allow the examination of some post-translational protein modifications which are functionally important for many proteins. It is possible to screen proteins systematically from reference maps to establish their identities.

To define protein-based gene expression analysis, the concept of the 'proteome' was recently proposed (Wilkins *et al.*, 1995; Wasinger *et al.*, 1995). A proteome is the entire PROTein complement expressed by a genome, or by a cell or tissue type. The concept of the proteome has some differences from that of the genome, as while there is only one definitive genome of an organism, the proteome is an entity which can change under different conditions, and can be dissimilar in different tissues of a single organism. A proteome nevertheless remains a direct product of a genome. Interestingly, the number of proteins in a proteome can exceed the number of genes present, as protein products expressed by alternative gene splicing or with different post-translational modifications are observed as separate molecules on a 2-D gel. As an extrapolation of the concept of the 'genome project', a 'proteome project' is research which seeks to identify and characterise the proteins present in a cell or tissue and define their patterns of expression.

Proteome projects present challenges of a similar magnitude to that of genome projects. Technically, the 2-D gel electrophoresis must be reproducible and of high resolution, allowing the separation and detection of the thousands of proteins in a cell. Low copy number proteins should be detectable. There should be computer gel image analysis systems that can qualitatively and quantitatively catalog the electrophoretically separated proteins, to form reference maps. A range of rapid and reliable techniques must be available for the identification and characterisation of proteins. As a consequence of a proteome project, protein databases must be assembled that contain reference information about proteins; such databases must be linked to genomic databases and protein reference maps. Databases should be widely accessible and easy to use.

Recently, there have been many changes in the techniques and resources available for the analysis of proteomes. It is the aim of this chapter to discuss the status of the areas outlined above, and to review briefly the progress of some current proteome projects.

Two-dimensional electrophoresis of proteomes

Two dimensional (2-D) gel electrophoresis involves the separation of proteins by their isoelectric point in the first dimension, then separation according to molecular weight by sodium dodecyl sulfate electrophoresis in the second dimension. Since first described (Klose, 1975; O'Farrell, 1975; Scheele, 1975), it has become the method of choice for the separation of complex mixtures of proteins, albeit with many modifications to the original techniques. 2-D electrophoresis forms the basis of proteome projects through separating proteins by their size and charge (Hochstrasser *et al.*,

Fig.
illu-
was
not
the
of L

19
pr
sin

2-1
A ;
ph-
en-

HEPG2 2D-PAGE MAP

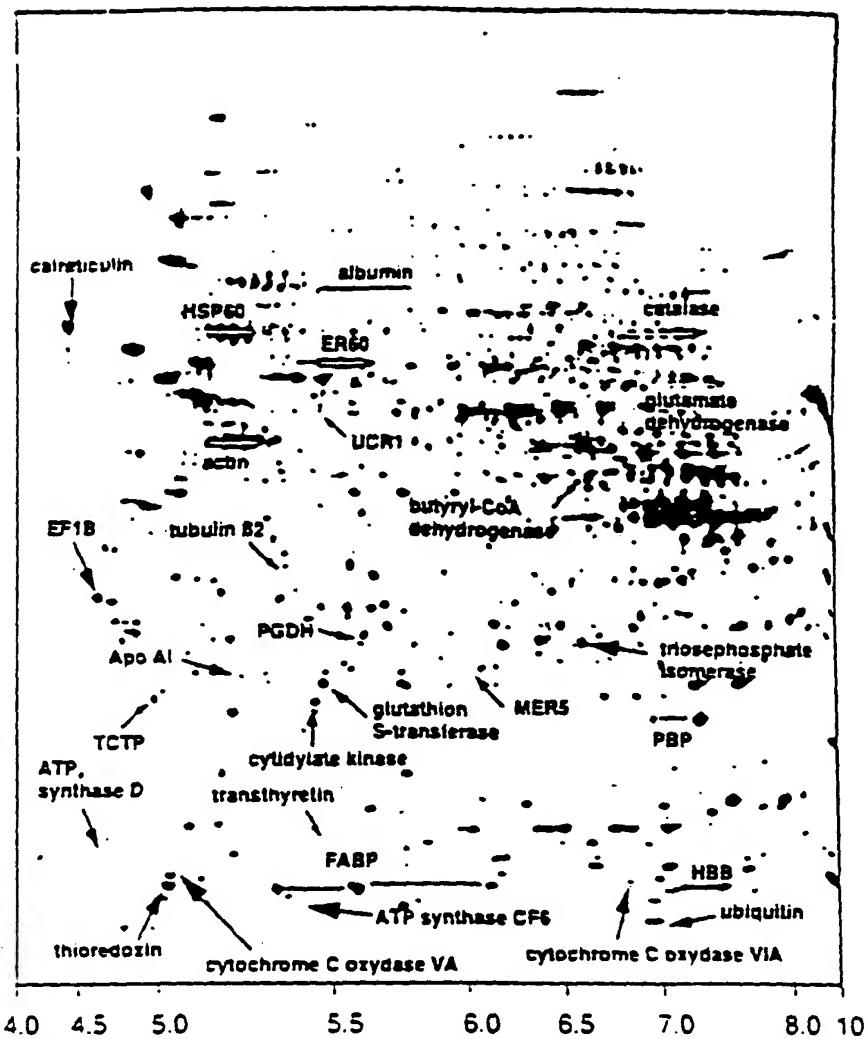


Figure 1. Two-dimensional gel electrophoresis map of a human hepatoblastoma-derived cell line, illustrating the very high resolution of the technique. The first dimensional separation (right to left of figure) was achieved using immobilised pH gradient electrophoresis of 4.0 to 10.0 units. The second dimension (top to bottom of figure) was SDS-PAGE using a 11%–14% acrylamide gradient, allowing separation in the molecular weight range 10–250 kDa. Proteins were visualised by silver staining. Arrows show proteins of known identity.

1992; Celis *et al.*, 1993; Gartels and Franzia, 1989; VanBogelen *et al.*, 1992). Current protocols can resolve two to three thousand proteins from a complex sample on a single gel (*Figure 1*).

2-D GEL RESOLUTION AND REPRODUCIBILITY

A primary challenge of separating complex mixtures of proteins by 2-D gel electrophoresis has been to achieve high resolution and reproducibility. High resolution ensures that a maximum of protein species are separated, and high reproducibility is

vital to allow comparison of gels from day to day and between research sites. These factors can be difficult to achieve.

CARRIER AMPHOLYTES are a common means of isoelectric focusing for the first dimension of 2-D electrophoresis. Gels are usually focused to equilibrium to separate proteins in the pI range 4 to 8, and run in a non-equilibrium mode (NEPHGE) to separate proteins of higher pI (7 to 11.5) (O'Farrell, 1975; O'Farrell, Goodman and O'Farrell, 1977). Unfortunately, the use of carrier ampholytes in the isoelectric focusing procedure is susceptible to 'cathode drift', whereby pH gradients established by prefocusing of ampholytes slowly change with time (Righetti and Drysdale, 1973). Carrier ampholyte pH gradients are also distorted by high salt concentration of samples (Bjellqvist *et al.*, 1982), and by high protein load (O'Farrell, 1975). A further limitation is that iso electric focusing gels, which are cast and subject to electrophoresis in narrow glass tubes, need to be extruded by mechanical means before application to the second dimension - a procedure that potentially distorts the gel. Nevertheless, many of the above shortcomings can be avoided by loading small amounts of ¹⁴C or ³⁵S radiolabelled samples (Garrels, 1989; Neidhardt *et al.*, 1989; Vandekerckhove *et al.*, 1990). High sensitivity detection is then achieved through use of fluorography or phosphorimaging plates (Bonner and Laskey, 1974; Johnston, Pickett and Barker, 1990; Patterson and Lamer, 1993). However, this approach is only practicable for organisms or tissues that can be radiolabelled.

An alternative technique, which is becoming the method of choice for the first dimension separation of proteins, involves isoelectric focusing in immobilized pH gradient (IPG) gels (Bjellqvist *et al.*, 1982; Görg, Postel and Günther, 1988; Righetti, 1990). Immobilized pH gradients are formed by the covalent coupling of the pH gradient into an acrylamide matrix, creating a gradient that is completely stable with time. IPG gels are usually poured onto a stiff backing film, which is mechanically strong and provides easy gel handling (Östergren, Eriksson and Bjellqvist, 1988). The major advantages of IPG separations are that they do not suffer from cathodic drift, they allow focusing of basic and very acidic proteins to equilibrium, pH gradients can be precisely tailored (linear, stepwise, sigmoidal), and that separations over a very narrow pH range are possible (0.05 pH units per cm) (Righetti, 1990; Bjellqvist *et al.*, 1982, 1993a; Sinha *et al.*, 1990; Görg *et al.*, 1988; Gelfi *et al.*, 1987; Günther *et al.*, 1988). However, it is not currently possible to use IPG gels to separate very basic proteins of isoelectric point greater than 10, although this is under development. Narrow pH range separations are useful to address problems of protein co-migration in complex samples, allowing 'zooming in' on regions of a gel (Figure 2). IPG gel strips are now commercially available, which begin to address the problems of intra- and inter-lab isoelectric focusing reproducibility.

There are two means of electrophoresis for the second dimension separation of proteins: vertical slab gels and horizontal ultrathin gels (Görg, Postel, and Günther, 1988). Both are usually SDS-containing gradient gels of approximately 11% to 15% acrylamide, which separate proteins in the molecular mass range of 10 - 150kD. A stacking gel is not usually used with slab gels, but is necessary when using horizontal gel setups (Görg, Postel and Günther, 1988). Comparisons have shown that there is little or no difference in the reproducibility of electrophoresis using either approach (Corbett *et al.*, 1994a), but commercially available vertical or horizontal precast gels will provide greater reproducibility for occasional users. For slab gel electrophoresis,

Fig
2 P
m
Th
wi
pla
sel
Pr

th
ca
de
je

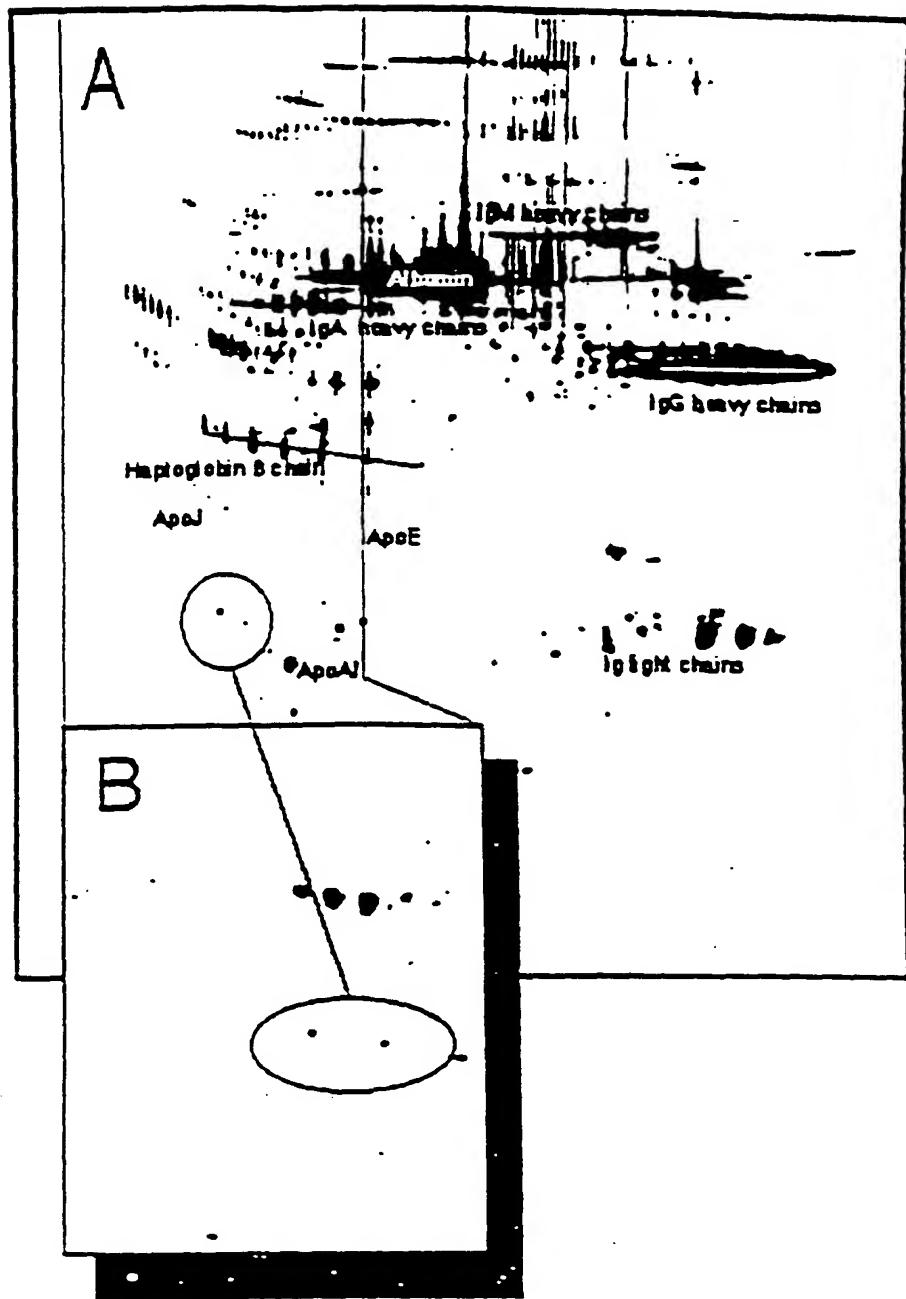


Figure 2. Two-dimensional gel electrophoresis allows 'zooming in' on areas of interest. Rings highlight 3 proteins common to each gel. (A) Wide pH range two-dimensional electrophoresis map of human plasma proteins. First dimension separation was achieved using an immobilised pH gradient of 3.5 to 10.0 units. The second dimension was SDS-PAGE. Actual gel size was 16cm x 20cm, and proteins were visualised with silver staining. (B) Narrow pH range electrophoresis was used to 'zoom in' on a small region of the plasma map. The first dimension used a narrow range immobilised pH gradient of 4.2 to 5.2 units, and second dimension was SDS-PAGE. Micropreparative loading was used, and the gel blotted to PVDF. Proteins were visualised with amido black. Actual blot size was 16cm x 20cm.

the use of piperazine diacrylyl as a gel crosslinker and the addition of thiosulfate in the catalyst system has been shown to give better resolution and higher sensitivity detection (Hochstrasser and Mertil, 1988; Hochstrasser, Patchornik and Mertil, 1988).

Notwithstanding the advances described above, there is an increasing demand to improve the reproducibility of 2-D electrophoresis to facilitate database construction and proteome studies. Harrington *et al.* (1993) explain that if a gel resolves 4000 protein spots, and there is 99.5% spot matching from gel to gel, this will produce 20 spot errors per gel. This amount of error, which might accumulate with each gel to gel comparison used in database construction, could produce an unacceptable degree of uncertainty in gel databases. To address these issues, partial automation of large 2-D gel separations has been undertaken (Nokihara, Morita and Kuriki, 1992; Harrington *et al.*, 1993). Although results are preliminary, spot to spot positional reproducibility in one study was found to be threefold improved over manual methods (Harrington *et al.*, 1993). It should be noted that small 2-D gel formats (50 x 43 mm) have been almost completely automated (Brewer *et al.*, 1986), although these are not generally used for database studies.

MICROPREPARATIVE 2-D GEL ELECTROPHORESIS

With the advent of affordable protein microcharacterisation techniques, including N-terminal microsequencing, amino acid analysis, peptide mass fingerprinting, phosphate analysis and monosaccharide compositional analysis, a new challenge for 2-D electrophoresis has been to maintain high resolution and reproducibility but to provide protein in sufficient quantities for chemical analysis (high nanogram to low microgram quantities of protein per spot). This becomes difficult to achieve with very complex samples such as whole bacterial cells, as the initial protein load is divided among 2000 to 4000 protein species. Two approaches are used for producing amounts of material that can be chemically characterised. The first method is to run multiple gels, collect and pool the spots of interest, and subject them to concentration (Ji *et al.*, 1994; Walsh *et al.*, 1995; Rasmussen *et al.*, 1992). In this approach, the concentration process must also act as a purification step to remove accumulated electroforetic contaminants such as glycine. A more elegant approach has been to exploit the high loading capacity of IPG isoelectric focusing. The high loading capacity of immobilised pH gradients was described early (Ek, Bjellqvist and Righetti, 1983), but has only recently been applied to 2-D electrophoresis (Hanash *et al.*, 1991; Bjellqvist *et al.*, 1993b). Up to 15 mg of protein can be applied to a single gel, yielding microgram quantities of hundreds of protein species. A further benefit of this approach is that proteins present in low abundance, which may not be visualised by lower protein loads, are more likely to be detected. The use of electroforetic or chromatographic prefractionation techniques (Hochstrasser *et al.*, 1991a; Harrington *et al.*, 1992), followed by high loading of narrow-range IPG separations (Bjellqvist *et al.*, 1993b) provides a likely solution to studies on proteins present in low abundance.

Methods of protein detection

There are many means for detecting proteins from 2-D gels. The method used will be dictated by factors including protein load on gel (analytical or preparative), the purpose of the gel (for protein quantitation or profiling and chemical characterisation), and the sensitivity required. The most common means of protein detection and their applications are shown in Table 1. Most detection methods have drawbacks, for

Table 1: Common stains for 2-D gels or blots and their applications.

Detection Method	Main applications	Unsuitable applications	Sensitivity	References
³⁵ S) Met or ¹⁴ C radiolabelling and fluorography or phosphorimaging	Cell lines, cultured organisms	Samples that cannot be labelled	20 ppm of radiolabel in a spot	Ganek and Franza, 1986; Lathum, Ganek and Sulter, 1993
³⁵ S) iodo-urea silver	Extremely high sensitivity gel staining	Preparative 2-D, PVDF or NC membranes	0.1 ng protein on spot or band of gel	Wallace and Saluz, 1992 a,b
Silver	Very high sensitivity gel staining, can be mono or polychromatic	Preparative 2-D, PVDF or NC membranes	1 ng protein on spot or band of gel	Rabilloud, 1992; Hochstrasser and Merril, 1998
Coomassie blue R-250	Staining of gels, staining of PVDF membranes before protein sequencing	Staining prior to direct mass determination from PVDF; amino acid analysis on PVDF; detection of some glycoproteins	30 ng protein on band or spot of gel	Sirupal <i>et al.</i> , 1993; Gharahdaghi <i>et al.</i> , 1992; Goldberg <i>et al.</i> , 1998; Sanchez <i>et al.</i> , 1992
Colloidal gold	Staining NC membranes, staining PVDF before direct MALDI-TOF	Gels	60 x higher than coomassie	Yamaguchi and Nakawa, 1988; Eckerskorn <i>et al.</i> , 1992; Sirupal <i>et al.</i> , 1993
Zinc imidazole	Reverse staining of gels or membranes; may be beneficial in MALDI-TOF of peptides	Where positive image is required	Higher than coomassie	Ortiz <i>et al.</i> , 1992; James <i>et al.</i> , 1993
Ponceau S and amido black	Staining higher protein loads on PVDF, for protein sequencing or amino acid analysis	Staining prior to direct mass determination from PVDF	1(X) ng protein on band or spot of gel	Sanchez <i>et al.</i> , 1992; Sirupal <i>et al.</i> , 1992; Wilkins <i>et al.</i> , 1995
India ink	Staining of membrane-bound proteins, staining PVDF before direct MALDI-TOF	Gel staining, not quantitative from protein to protein	1–10 ng	Li <i>et al.</i> , 1989; Hughes, Mack and Hamparian, 1988; Sirupal <i>et al.</i> , 1993
Stain-all	Staining to detect glycoproteins or Ca ²⁺ binding proteins	General gel staining	1(X) ng protein on band or spot of gel	Campbell, MacLennan and Jorgenson, 1983; Goldberg <i>et al.</i> , 1998

PVDF = polyvinylidene difluoride, NC = nitrocellulose, MALDI-TOF = matrix assisted laser desorption ionisation time of flight mass spectrometry.

example, some glycoproteins are not stained by coomassie blue (Goldberg *et al.*, 1988), and many organic dyes are unsuitable for protein detection on PVDF if samples are to be used for direct matrix-assisted laser desorption ionisation mass spectrometry (Sirupal *et al.*, 1994).

Although most means of protein detection give some indication of the quantities of protein present, in general they cannot be used for global quantitation. This is because

no protein stain is able consistently to detect proteins over a wide range of concentrations, isoelectric points and amino acid compositions, and with a variety of post-translational modifications (Goldberg *et al.*, 1985; Li *et al.*, 1989). Furthermore, there are large differences in staining pattern when identical gels or blots are subjected to different stains, including amido black, imidazole zinc, india ink,ponceau S, colloidal gold, or coomassie blue (Tovey, Ford and Baldo, 1987; Ortiz *et al.*, 1992). The most common means of quantitating large numbers of protein, in a 2-D gel involves the radiolabelling of protein samples prior to electrophoresis, and protein quantitation based on fluorography and image analysis or liquid scintillation counting (Garrett, 1989; Celis and Olsen, 1994). However, proteins which do not contain methionine cannot be detected if only [³⁵S] methionine is used for labelling. Amino acid analysis of protein spots visualised by other techniques presents a likely means of protein quantitation for the future.

BLOTTING OF PROTEINS TO MEMBRANES

Electrophoretic blotting of proteins from two-dimensional polyacrylamide gels to membranes presents many options for protein identification and microcharacterisation which are not possible when proteins remain in gels. For example, when proteins are blotted to polyvinylidene difluoride (PVDF) membranes, they can be identified by N-terminal sequencing, amino acid analysis, or immunoblotting, or they may be subjected to endoproteinase digestion, monosaccharide analysis, phosphate analysis, or direct matrix-assisted laser desorption ionisation mass spectrometry (Matsudaira, 1987; Wilkins *et al.*, 1995; Jungblut *et al.*, 1994; Sutton *et al.*, 1995; Rasmussen *et al.*, 1994; Weitzhandler *et al.*, 1993; Murthy and Iqbal, 1991; Eckerskorn *et al.*, 1992). It is possible to combine some of these procedures on a single protein spot on a PVDF membrane (Packer *et al.*, 1995; Wilkins *et al.*, submitted; Weitzhandler *et al.*, 1993). This is useful when minimal amounts of protein are available for analysis. These techniques will be explored in detail later in this review. Notwithstanding the above, there are some disadvantages associated with blotting of proteins to membranes. There is always loss of sample during blotting procedure (Eckerskorn and Lutzpeich, 1993), and common protein detection methods are less sensitive or not applicable to membranes (Table 1), presenting difficulties for the analysis of low abundance proteins. Detailed discussion of the merits of available membranes and common blotting techniques can be found elsewhere (Eckerskorn and Lutzpeich, 1993; Sirupai *et al.*, 1994; Patterson, 1994).

2-D gel analysis, documentation, and proteome databases

Following protein electrophoresis and detection, detailed analysis of gel images is undertaken with computer systems. For proteome projects, the aim of this analysis is to catalogue all spots from the 2-D gel in a qualitative and if possible quantitative manner, so as to define the number of proteins present and their levels of expression. Reference gel images, constructed from one or more gels, form the basis of two-dimensional gel databases. These databases also contain protein spot identities and

details of their post-translational modifications. 2-D gel databases are beginning to be linked to or integrated with comprehensive protein and nucleic acid databases (Neidhardt *et al.*, 1989; Simpson *et al.*, 1992; Appel *et al.*, 1994), and 'organism' databases, containing DNA sequence data, chromosomal map locations, reference 2-D gels and protein functional information for an organism, are becoming established as genome and proteome projects progress (VanBogelen *et al.*, 1992; Yeast Protein Database cited in Garrels *et al.*, 1994).

GEL IMAGE ANALYSIS AND REFERENCE GELS

After 2-D electrophoresis and protein visualisation by staining, fluorography or phosphorimaging, images of gels are digitised for computer analysis by an image scanner, laser densitometer, or charge-coupled device (CCD) camera (Garrels, 1989; Celis *et al.*, 1990a; Urwin and Jackson, 1993). All systems digitise gels with a resolution of 100–200 mm, and can detect a wide range of densities or shading (256 or more 'grey scales'). Following this, gel images are subjected to a series of manipulations to remove vertical and horizontal streaking and background haze, to detect spot positions and boundaries, and to calculate spot intensity (*Figure 3*). A standard spot (SSP) number, containing vertical and horizontal positional information, is assigned to each detected spot and becomes the protein's reference number. *Table 2* lists some notable software packages which process 2-D gel images.

Table 2: Some Software Packages for the Analysis of Gel Images.

Gel Image Analysis System	References*
ELSIE 4 & 5	Olsen and Miller, 1988; Wirth <i>et al.</i> , 1991; Wirth <i>et al.</i> , 1993
GELLAB I & II	Wu, Lemkin and Lipson, 1993; Lemkin, Wu and Lipson, 1993; Myrick <i>et al.</i> , 1993.
MELANIE I & II	Appel, <i>et al.</i> , 1991; Hochstrasser <i>et al.</i> , 1991b
QUEST I & II and PDQUEST	Garrels, 1989; Monardini <i>et al.</i> , 1992; Holt <i>et al.</i> , 1992; Celis <i>et al.</i> , 1990a,b
TYCHO & KEPLAR	Anderson <i>et al.</i> , 1984; Richardson, Horn and Anderson, 1982

* These references are not exhaustive, they include some references of use as well as authors of the system.

As there are difficulties in the electrophoresis of samples with 100% reproducibility, reference gel images are often constructed from many gels of the same sample (Garrels and Franz, 1989; Neidhardt *et al.*, 1989). Since this involves the matching of 2000 to 4000 proteins from one gel to another, it presents a considerable challenge to image analysis systems. Matching of gels is usually initiated by an operator, who manually designates approximately 50 or so prominent spots as 'landmarks' on gels to be cross-matched. Proteins which match are then established around landmarks, using computer-based vector algorithms to extend the matching over the entire gel. Close to 100% of spots from complex samples can be matched by these methods, although different degrees of operator intervention may be required (Olsen and Miller, 1988; Lemkin and Lester, 1989; Garrels, 1989; Myrick *et al.*, 1993).

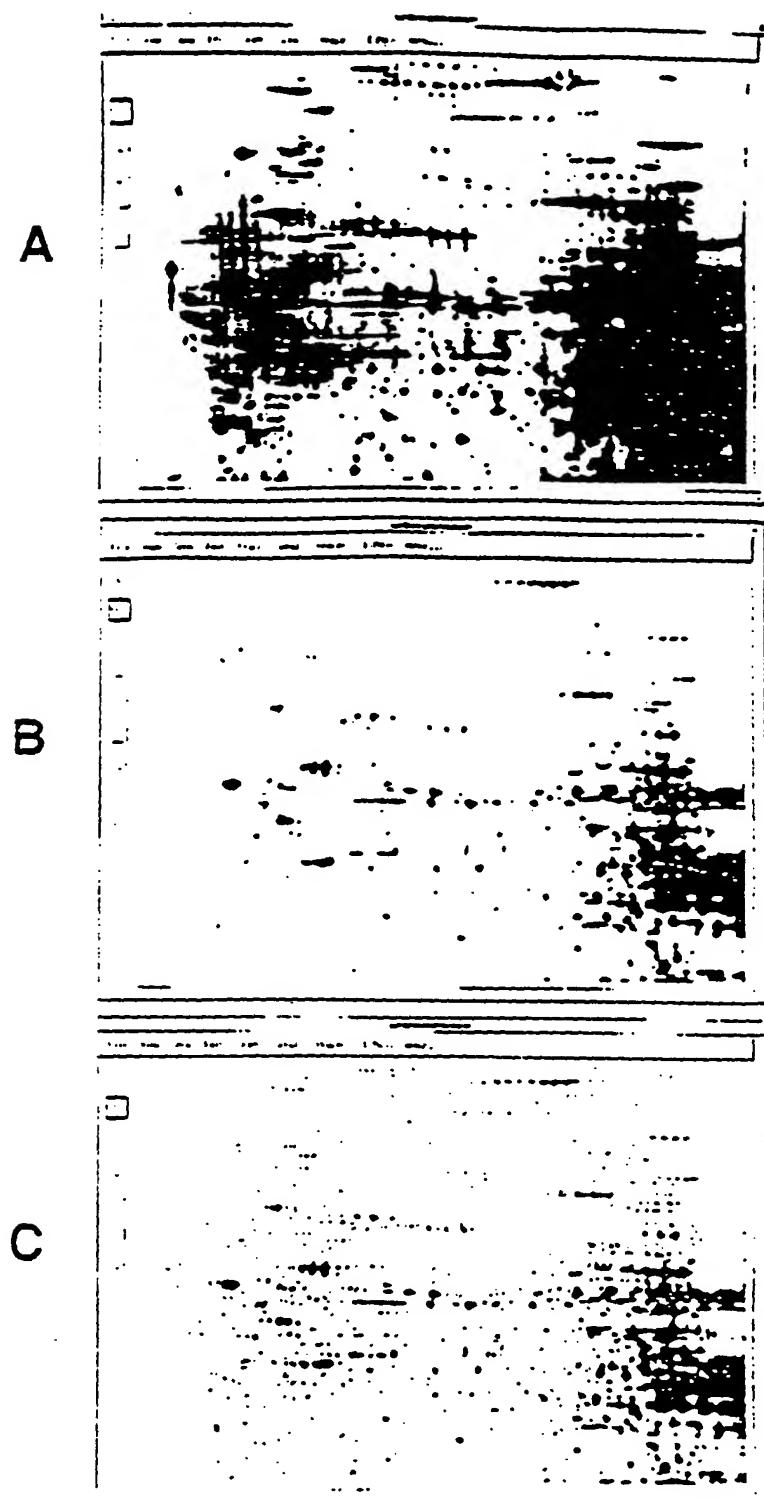


Figure 3. Computer processing of gel images. Shown is a wide pI range 2-D separation of human liver proteins, processed by Melanie software (Appel *et al.*, 1991). (A) Original gel image as captured by laser densitometer. (B) Gel image after processing to remove streaking and background. (C) Outline definition of all spots on the gel.

SPOT C
A major
separatio
to dete
positio
of man
person
minute
scintill
by ref
protein
et al.
Garrel
be no
Limit:
not ac
only e
an alti
Myric
1992.
Wh
and ni
protei
their a
transf
regula
synthe
(Lath.

CALCULATION OF PROTEIN ISOELCTRIC POINT AND MOLECULAR WEIGHT

Estimation of the isoelectric point (*pI*) and molecular weight (MW) of proteins from 2-D gel provides fundamental parameters for each protein, which are also of use during identification procedures (see following section). The *pI* and MW of proteins are recorded in 2-D gel databases. Accurate estimations of protein *pI* and MW can be obtained by using 20 or more known proteins on a reference map to construct standard curves of *pI* and molecular weight, which are then used to calculate estimated *pI* and MW of unknown proteins (Neidhardt *et al.*, 1989; Garrels and Franz, 1989; Vandekerckhove, Hutton and Neidhardt, 1990; Anderson and Anderson, 1991; Anderson *et al.*, 1991; Latham *et al.*, 1992). Alternatively, the MW of individual proteins blotted to PVDF can be determined very accurately by direct mass spectrometry (Eckerkorn *et al.*, 1992). Where immobilised pH gradients are used, the focusing position of proteins allows their *pI* to be measured within 0.15 units of that calculated from the amino acid sequence (Bjellqvist *et al.*, 1993c). It must be noted, however, that proteins carrying post-translational modifications may migrate to unexpected *pI* or MW positions during electrophoresis (Packer *et al.*, 1995).

SPOT QUANTITATION AND EXPRESSION ANALYSIS

A major challenge faced in proteome projects is the quantitative analysis of proteins separated by 2-D electrophoresis. The most accurate means of protein quantitation is to determine chemically the amount of each protein present by amino acid compositional analysis. However, the current method of choice for quantitative analysis of many proteins is to radiolabel samples with [³⁵S] methionine or [¹⁴C] amino acids, perform the 2-D electrophoresis, and measure protein levels in disintegrations per minute (dpm) or units of optical density. Quantitation is achieved either by liquid scintillation counting, or by gel image analysis where spot densities are quantitated by reference to gel calibration strips containing known amounts of radiolabelled protein or against the integrated optical density of all spots visualised (Vandekerckhove *et al.*, 1990; Celis *et al.*, 1990b; Celis and Olsen, 1994; Garrels, 1989; Latham, Garrels and Solter, 1993; Fey *et al.*, 1994). All approaches effectively allow spots to be normalised against the total disintegrations per minute loaded onto the gel. Limitations that remain with radiolabelling methods are that absolute quantitation is not achieved because all proteins have varying amounts of any amino acid, and that only easily labelled samples can be investigated. Quantitative silver staining presents an alternative (Giometti *et al.*, 1991; Harrington *et al.*, 1992; Rodriguez *et al.*, 1993; Myrick *et al.*, 1993), which when undertaken with [³⁵S] thiourea (Wallace and Saluz, 1992 a,b) is of extremely high sensitivity.

When protein spots from samples prepared under different conditions are quantitated and matched from gel to gel, it becomes possible to examine changes and patterns in protein expression. Large scale investigation of up- and down-regulation of proteins, their appearance and disappearance, can be undertaken. For example, simian virus 40 transformed human keratinocytes were shown to have 177 up-regulated and 58 down-regulated proteins compared to normal keratinocytes (Celis and Olsen, 1994); detailed synthesis profiles of 1200 proteins have been established in 1 to 4 cell mouse embryos (Latham *et al.*, 1991, 1992); and 4 proteins out of 1971 were found to be markers for

cadmium toxicity in urinary proteins (Myrick *et al.*, 1993). Complex global changes in protein expression as a result of gene disruptions have also been investigated (S. Fey and P. Moen-Larsen, Personal communication). Impressively, large gel sets showing protein expression under different conditions can be globally investigated using statistical methods that find groups of related objects within a set. For example, the REF52 rat cell line database, consisting of 79 gels from 12 experimental groups where each gel contains quantitative data for 1600 cross-matched proteins, has been analysed by cluster analysis (Garrels *et al.*, 1990). This revealed clusters of proteins that, for example, were induced or repressed similarly under simian virus 40 or adenovirus transformation, suggesting a common mechanism. Protein groups that were induced or repressed during culture growth to confluence were also found. It is obvious that the potential for investigation of cellular control mechanisms by these approaches is immense. It is equally clear that investigations of gene expression of this scale are currently technically impossible using nucleic-acid based techniques.

Table 3: Some proteome databases and their special features

Proteome database	Special features	References	
<i>E. coli</i> gene-protein database	Gel spots linked with GenBank and Kohara clones; quantitative spot measurements under different growth conditions	VanBogelen and Neidhardt, 1991. VanBogelen <i>et al.</i> , 1992	One SWISS 1994; feature 2DPAGE
Human heart database	Identification of disease markers; two separate databases have been established	Baker <i>et al.</i> , 1992 Correll <i>et al.</i> , 1994b Jungblut <i>et al.</i> , 1993	Table 4 All three expressed
Human keratinocyte database	Extensive identifications; quantitative spot measurements of transformed cells; identification of disease markers	Celis <i>et al.</i> , 1990a Celis <i>et al.</i> , 1993 Celis and Olsen, 1994	Informa-
Mouse embryo database	Quantitative spot measurements through 1 to 4 cell stage	Latham <i>et al.</i> , 1991 Latham <i>et al.</i> , 1992	
Mouse liver database (Argonne Protein Mapping Group)	Documents changes due to exposure to ionizing radiation and toxic chemicals	Gionnet, Taylor and Tollaksen, 1992	Annals
Rat liver epithelial database	Detailed subcellular fractionation studies	Wirth <i>et al.</i> , 1991 Wirth <i>et al.</i> , 1993	
Rat liver database	Extensive studies on regulation of proteins by drugs and toxic agents	Anderson and Anderson, 1991. Anderson <i>et al.</i> , 1992. Richardson, Horn and Anderson, 1992	
REF 52 rat cell line database	Accessible via World Wide Web; quantitative spot measurements under different conditions	Garrels and Franza, 1989 Bouelli <i>et al.</i> , 1992	Chick- Refere- Datab.
SWISS-2DPAGE containing human reference maps	Accessible via World Wide Web; completely integrated with SWISS-PROT and SWISS-3DIMAGE	Appel <i>et al.</i> , 1992 Hochstrasser <i>et al.</i> , 1992 Hughes <i>et al.</i> , 1993 Gobet <i>et al.</i> , 1993	
Yeast Protein Database (YPD) and Yeast Electrophoretic Protein Database (YEPD)	Completely crossreferenced organism database. YPD has extensive information on over 3500 proteins; YEPD has many identifications	Garrels <i>et al.</i> , 1994	Other

FEATURES OF PROTEOME DATABASES

Proteome projects rely heavily on computer databases to store information about all proteins expressed by an organism. 'Proteome databases' should contain detailed information of proteins already characterised elsewhere, as well as protein data from 2-D gels such as apparent pI and MW, expression level under different conditions, subcellular localisation, and information on post-translational modifications. Images of reference 2-D gels, showing protein SSP numbers and protein identifications, should also be included. Ideally, proteome databases should be accessible with Macintosh or IBM personal computers and easy to use. Some proteome databases and the areas they cover are listed in *Table 3*. Databases range from collections of annotated gels to large databases of images integrated with protein and nucleic acid sequence banks.

One example of an integrated proteome database is the suite of SWISS-PROT, SWISS-2DPAGE and SWISS-3DIMAGE databases (Appel *et al.*, 1993; Appel *et al.*, 1994; Appel, Bairoch and Hochstrasser, 1994; Bairoch and Boeckmann, 1994). The features of these three databases are listed in *Table 4*. SWISS-PROT, SWISS-2DPAGE and SWISS-3DIMAGE are accessible through the World Wide Web

Table 4: The SWISS-PROT, SWISS-2DPAGE and SWISS-3DIMAGE suite of crosslinked databases. All three databases are accessible through the World Wide Web, at URL address: <http://expasy.hcuge.ch/>

	SWISS-PROT	SWISS-2DPAGE	SWISS-3DIMAGE
Information	Text entries of sequence data; Citation information: taxonomic data, 38, 303 entries in Release 29	2-D gel images of: human liver, plasma, HepG2, HepG2 secreted proteins, red blood cell, lymphoma, cerebrospinal fluid, macrophage like cell line, erythroleukemia cell, platelet	Collection of 330 3-D images of proteins
Annotations	Protein function, Post translational modifications, Domains, Secondary structure, Quaternary structure, Diseases associated with protein, Sequence conflicts	Gel images where protein is found, How protein identified, Protein pI and MW, protein number, normal and pathological variants	All annotation is available in SWISS-PROT
Cross-referenced Databases	SWISS-2DPAGE, SWISS-3DIMAGE, EMBL, PIR, PDB, OMIM, PROSITE, Medline, Flybase, GCRDb, MaizeDB, WonnPep, DictpyDB	SWISS-PROT and all other databases accessible through SWISS-PROT	SWISS-PROT and all other databases accessible through SWISS-PROT
Other Features	Navigation in other SWISS databases achieved by selecting entries with computer mouse	Gel images show position of identified proteins, or region of gel where protein should appear	Mono and stereo images available, Images can be transferred to local computer image viewing programs

(Berners-Lee *et al.*, 1992), allowing any computer connected to the internet to access the stored information and images. Navigation within and between the three databases is seamless, as all potential crosslinks are highlighted as hypertext on the display and can be selected with a computer mouse. From these databases, detailed information about a protein, including amino acid sequence and known post-translational modifications, can be obtained, the precise protein spot it corresponds to on a reference gel image can be viewed if known, and the 3-D structure of the molecule can be seen if available. References to nucleic acid and other databases are also given to provide access to information stored elsewhere.

'Organism' databases, containing detailed protein and nucleic acid information about a species, are becoming common as genome and proteome projects progress. These differ from nucleic acid or protein sequence databases like GenBank or SWISS-PROT because they are image based, and contain information about chromosomal map positions, transcription of genes, and protein expression patterns. The *Escherichia coli* gene-protein database (VanBogelen, Hutton and Neidhardt, 1990; VanBogelen and Neidhardt, 1991; VanBogelen *et al.*, 1992), known as the ECO2DBASE, is one example. It contains gene and protein names, 2-D gel spot information (including pI and MW estimates, and spot identification), genetic information (GenBank or EMBL codes, chromosomal location, location on Kohara clones (Kohara, Akiyama, and Isono, 1987), transcription direction of genes), and protein regulatory information (level of protein expression under different growth regimes, member of regulon or operon). All entries in the ECO2DBASE are also cross-referenced to the SWISS-PROT database (Bairoch and Boeckmann, 1994). It is anticipated that organism databases will soon become a standard means of storing all available information about a particular species. However there is currently no consistent manner in which organism databases are assembled, which may hamper comparisons in the future.

Identification and characterisation of proteins from 2-D gels

The number of proteins identified on a 2-D reference map determines its usefulness as a research and reference tool. As most reference maps have only a small proportion of proteins identified, a major aim of current proteome projects is to screen many proteins from 2-D maps, in order to define them as 'known' in current nucleic acid and protein databases, or as 'unknown'. Protein identification assists in confirmation of DNA open reading frames, and provides focus for DNA sequencing projects and protein characterisation efforts by pointing to proteins that are novel. Since there may be 3000–4000 proteins from a single 2-D map that require identification, the challenge in protein screening is to identify proteins quickly, with a minimum of cost and effort.

Traditionally, proteins from 2-D gels have been identified by techniques such as immunoblotting, N-terminal microsequencing, internal peptide sequencing, comigration of unknown proteins with known proteins, or by overexpression of homologous genes of interest in the organism under study (Matsudaira, 1987; Rosenfeld *et al.*, 1992; VanBogelen *et al.*, 1992; Celis *et al.*, 1993; Honore *et al.*, 1993; Garrels *et al.*, 1994). Whilst these techniques are powerful identification tools, they are too expensive or time and labour intensive to use in mass screening programs. A hierarchical approach to mass protein identification has been recently suggested as an

Table 5: Hierarchical analysis for mass screening of 2-D separated proteins blotted to membrane. Rapid and inexpensive techniques are used as a first step in protein identification, and slower, more expensive techniques are then used if necessary. Table modified from Wasinger *et al.*, 1995.

Order	Identification technique	Reference
1	Amino acid analysis	Jungblut <i>et al.</i> , 1992; Shaw, 1993;
2	Amino acid analysis with N-terminal sequence tag	Hohmann, Houthaeve and Sander, 1992;
3	Peptide-mass fingerprinting	Jungblut <i>et al.</i> , 1994; Wilkins <i>et al.</i> , 1995; Wilkins <i>et al.</i> , submitted
4	Combination of amino acid analysis and peptide mass-fingerprinting	Henzel <i>et al.</i> , 1993; Pappin, Hiorup and Bleasby, 1993; James <i>et al.</i> , 1993;
5	Mass spectrometry sequence tag	Mann, Hiorup and Roepstorff, 1993;
6	Extensive N-terminal Edman microsequencing	Yates <i>et al.</i> , 1993; Mann <i>et al.</i> , 1992;
7	Internal peptide Edman microsequencing	Sutton <i>et al.</i> , 1995
8	Microsequencing by mass spectrometry (electrospray ionisation, post-source decay MALDI-TOF)	Cordwell <i>et al.</i> , 1995; Wasinger <i>et al.</i> , 1995;
9	Ladder sequencing	Mann and Wilms, 1993; Matsudaira, 1987; Rosenfeld <i>et al.</i> , 1992; Hellman <i>et al.</i> , 1993; Johnson and Walsh, 1992; Bartle-Jones <i>et al.</i> , 1992

alternative to traditional approaches (*Table 5*; Wasinger *et al.*, 1995). This involves the use of rapid and cheap identification tools such as amino acid analysis and peptide mass-fingerprinting as first steps in protein identification, followed by the use of slower, more expensive and time consuming identification procedures if necessary. In the construction of this hierarchy the analysis time, cost per sample and the complexity of the data created has been considered, as whilst some techniques require little machine time per sample, the analysis of data can be quite involved and time consuming. Amino acid analysis and peptide mass-fingerprinting based identification techniques in the hierarchy are discussed in detail below. For review of other protein identification techniques in *Table 5*, see Patterson (1994) and Mann (1995).

PROTEIN IDENTIFICATION BY AMINO ACID COMPOSITION

There has been a revival of interest in the use of amino acid composition for identification of proteins from 2-D gels after early work by Eckerskorn *et al.* (1988). This technique uses a protein's idiosyncratic amino acid composition profile in order to identify it by comparison with theoretical compositions of proteins in databases. The amino acid composition of proteins can be determined by differential metabolic radiolabelling and quantitative autoradiography after 2-D electrophoresis (Garrels *et al.*, 1994; Frey *et al.*, 1994), or by acid hydrolysis of membrane-blotted proteins and chromatographic analysis of the resulting amino acid mixture (Eckerskorn *et al.*, 1988; Tous *et al.*, 1989; Gharahdaghi *et al.*, 1992; Jungblut *et al.*, 1992; Wilkins *et al.*, 1995). As differential metabolic labelling experiments require X-ray film or phosphor-image plate exposures of up to 140 days, and can only be undertaken with easily radiolabelled samples, the technique is not as rapid or widely applicable as chromato-

Spot ECOLI-BLM

Composition:

Axx:	13.2	Glx:	10.4	Ser:	5.7	His:	0.7
Gly:	5.4	Cys:	3.6	Ala:	6.7	Pro:	7.9
Tyr:	1.3	Arg:	5.0	Val:	8.0	Met:	0.3
Ile:	5.9	Leu:	6.0	Phe:	13.3	Lys:	4.4

pI estimate: 6.89 Range searched: (6.64, 7.14)
 Mw estimate: 16800 Range searched: (13660, 20160)

Closest SWISS-PROT entries for the species ECOLI matched by AA composition:

Rank	Score	Protein	pI	Mw	Description
1	24	PYKE_ECOLI	6.84	16989	ASPARTATE CARBOXYLTRANSFERASE
2	39	CDAA_ECOLI	6.32	36359	PANTOTHENATE KINASE (EC 2.7.1.33)
3	40	META_ECOLI	5.06	35713	HOMOSERINE O-SUCINYLTRANSFERASE
4	42	CADC_ECOLI	5.52	57812	TRANSCRIPTIONAL ACTIVATOR CADC.
5	43	KLYC_ECOLI	8.36	19769	HEMOLYSIN C. PLASMID.

Closest SWISS-PROT entries for ECOLI with pI and Mw values in specified range:

Rank	Score	Protein	pI	Mw	Description
1	24	PYKE_ECOLI	6.84	16989	ASPARTATE CARBOXYLTRANSFERASE
2	102	TRAJ_ECOLI	6.73	17921	TRAJ PROTEIN.
3	112	YAJG_ECOLI	6.79	19028	HYPOTHETICAL LIPOPROTEIN YAJG.
4	140	YFJB_ECOLI	6.83	14945	HYPOTHETICAL 14.9 KD PROTEIN IN GRPE
5	142	YAHB_ECOLI	7.06	14726	HYPOTHETICAL PROTEIN IN BATT 3' REGION

Figure 4. Computer printout from ExPASy server where the empirical amino acid composition, estimated pI and MW of a protein from a 2-D reference map of *E. coli* were matched against all entries in SWISS-PROT for *E. coli*. The correct identification, aspartate carboxyltransferase, is shown in bold. Low scores indicate a good match. Note how matching within a defined pI and MW range (lower set of proteins) has greatly increased the score difference between the first and second ranking proteins. This score difference gives high confidence in the identification, and is only observed where the top ranking protein is the correct identification (Wilkins *et al.*, 1995).

graphy-based analysis. Proteins blotted to PVDF membranes can be hydrolysed in 1 h at 155°C, amino acids extracted in a single brief step, and each sample automatically derivatized and separated by chromatography in under 40 minutes (Wilkins *et al.*, 1995; Ou *et al.*, 1995). In this manner, one operator can routinely analyse 100 proteins per week on one HPLC unit. This technology lends itself to automation, and it is anticipated that instruments with even greater sample throughput will be developed. When proteins have been prepared by micropreparative 2-D electrophoresis (Hanash *et al.*, 1991; Bjellqvist *et al.*, 1993b), blotted to a PVDF membrane and stained with amido black, any visible protein spot is of sufficient quantity for amino acid analysis (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995; Wilkins *et al.*, 1995).

After the amino acid composition of a protein has been determined, computer programs are used to match it against the calculated compositions of proteins in databases (Eckerskorn *et al.*, 1988; Sibbald, Sommersfeldt and Argos, 1991; Jungblut *et al.*, 1992; Shaw, 1993; Hobohm, Houthaeve and Sander, 1994; Wilkins *et al.*, 1995). Matching is usually done with only 15 or 16 amino acids, as cysteine and

tryptophan are not included in the scoring system. The confidence in the identification is proportional to the rank of the top scoring protein.

The confidence in the identification is proportional to the rank of the top scoring protein. The confidence in the identification is proportional to the rank of the top scoring protein.

When the confidence in the identification is high, the top scoring protein is likely to be the correct identification.

Spot ECCL1-ACG

Composition:

Asx:	9.4	Glx:	10.8	Ser:	6.1	His:	2.7
Gly:	12.2	Thr:	2.8	Ala:	11.9	Pro:	3.2
Tyr:	6.0	Arg:	3.7	Val:	9.5	Met:	0.6
Ile:	5.0	Leu:	8.2	Phe:	3.2	Lys:	4.9

pI estimate: 5.99 Range searched: (5.76, 6.24)
Mw estimate: 45000 Range searched: (36000, 54000)

Closest SWISS-PROT entries for ECCL1 with pI and Mw values in specified range:

Rank	Score	Protein	pI	Mw	N-terminal Seq.
1	21	GLTA_ECOLI	6.03	45316	M I K R E
2	32	YGBB_ECOLI	5.86	36502	M S M I K
3	38	GAAT_ECOLI	5.78	45774	M S N S K
4	44	YIMS_ECOLI	5.86	48018	M R I K Y
5	45	DHE4_ECOLI	5.98	48581	M D Q T Y
6	46	ARGB_ECOLI	5.79	43765	M A I E Q
7	46	MJPA_ECOLI	5.78	37851	M N H S L
8	47	GLMK_ECOLI	5.98	49162	M L N R A
9	47	ACKA_ECOLI	5.85	43290	M S S K L
10	50	YGBH_ECOLI	6.01	37064	M E S K E

Figure 5. A PVDF protein spot from an *E. coli* 2-D reference map was sequenced for 4 cycles, and the same sample then subject to amino acid analysis. The N-terminal sequence was M L K R. When the amino acid composition of the spot, as well as estimated pI and MW, were matched against all entries in SWISS-PROT for *E. coli*, the above list of best matches was produced. N-terminal sequences are from SWISS-PROT for those entries. The top ranking identification of serine hydroxymethyltransferase (bold) did not show a large score difference between the first and second ranking proteins, giving little confidence in this being the correct protein identification. However, the sequence tag (M L K R) confirmed the identity of the protein as serine hydroxymethyl transferase.

tryptophan are destroyed during hydrolysis, asparagine and glutamine are deamidated to their corresponding acids, and proline is not quantitated in some analysis systems. The computer programs produce a list of best matching proteins, which are ranked by a score that indicates the match quality. Some programs allow matching to be restricted to specific 'windows' of MW and pI (Hobohm, Houthaeve and Sander, 1994; Wilkins *et al.*, 1995), and to protein database entries for one species (Jungblut *et al.*, 1992; Wilkins *et al.*, 1995). The use of such restrictions increases the power of matching. An example of protein identification by amino acid composition is shown in Figure 4. To date, amino acid composition has been used to identify proteins from reference maps of *Spiroplasma melliferum*, *Mycoplasma genitalium*, *E. coli*, *Saccharomyces cerevisiae*, *Dicyostelium discoideum*, human sera, human heart, human lymphocyte, and mouse brain (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995; Wilkins *et al.*, 1995; Jungblut *et al.*, 1992, 1994; Garrels *et al.*, 1994; Frey *et al.*, 1994).

PROTEIN IDENTIFICATION BY AMINO ACID COMPOSITION AND N-TERMINAL SEQUENCE TAG

When samples from 2-D gels are not unambiguously identified by amino acid

composition, pI and MW, often the correct identification of that protein is among the top rankings of the list (Hobohm, Houthaeve and Sander, 1994; Cordwell *et al.*, 1995; Wilkins *et al.*, 1995). Taking advantage of this observation, we have used the mass-spectrometry 'sequence tag' concept (Mann and Wilkins, 1994) in developing a combined Edman degradation and amino acid analysis approach to protein identification (Wilkins *et al.*, submitted). This involves the N-terminal sequencing of PVDF-blotted protein by Edman degradation for 3 or 4 cycles to create a 'sequence tag', following which the same sample is used for amino acid analysis. As only a few amino acids are removed from the protein, its composition is not significantly altered. Furthermore, since only a small amount of protein sequence is required, fast but low repetitive yield Edman degradation cycles can be used. Modifications to current procedures should allow 3 cycles to be completed in 1 h, thereby allowing the screening of 100 or more proteins per week on one automated, multi-cartridge sequenator. Amino acid composition, pI and MW of proteins are matched against databases as described above, and N-terminal sequences of best matching proteins are checked with the 'sequence tag' to confirm the protein identity (Figure 5). This technique will be less useful when proteins are N-terminally blocked, but as only a few N-terminal amino acids are susceptible to the acetyl, formyl, or pyroglutamyl modifications that cause blockage, this may itself provide useful information for sequence tag identification. A strength of N-terminal sequence tag and amino acid composition protein identification is that data generated are quickly and easily interpreted.

PROTEIN IDENTIFICATION BY PEPTIDE MASS FINGERPRINTING

Techniques for the identification of proteins by peptide mass fingerprinting have recently been described (Henzel *et al.*, 1993; Pappin, Hojrup and Bleasby, 1993; James *et al.*, 1993; Mann, Hojrup and Roepstorff, 1993; Yates *et al.*, 1993; Monz *et al.*, 1994; Sutton *et al.*, 1995). This involves the generation of peptides from proteins using residue-specific enzymes, the determination of peptide masses, and the matching of these masses against theoretical peptide libraries generated from protein sequence databases. As proteins have different amino acid sequences, their peptides should produce characteristic 'fingerprints'.

The first step of peptide mass fingerprinting is protein digestion. Proteins within the gel matrix or bound to PVDF can be enzymatically digested *in situ*, although *in situ* gel digests are reported to produce more enzyme autodigestion products, which complicate subsequent peptide mass analysis (James *et al.*, 1993; Rasmussen *et al.*, 1994; Monz *et al.*, 1994). The enzyme of choice for digestion is currently trypsin (of modified sequencing grade), but other enzymes (Lys-C or *S. aureus* V8 protease) have also been used (Pappin, Hojrup and Bleasby, 1993). To maximise the number of peptides obtained, it is desirable for protein samples to be reduced and alkylated prior to digestion (Monz *et al.*, 1994; Henzel *et al.*, 1993). This ensures that all disulphide bonds of the protein are broken, and produces protein conformations that are more amenable to digestion. Surprisingly, chemical digestion methods such as cyanogen bromide (methionine specific), formic acid (aspartic acid specific), and 2-(2-nitrophenylsulfonyl)-3-methyl-3'-bromoindolenine (tryptophan specific) have not been explored as means of peptide production for mass fingerprinting, even though they are rapid and may circumvent some problems associated with enzyme digestions.

(Nikodem and Freco, 1979; Crimmins *et al.*, 1990; Vanfleteren *et al.*, 1992).

After proteins are digested, peptide masses are determined by mass spectrometry. Direct analysis of peptide mixtures can be achieved by electrospray ionisation mass spectrometry, plasma desorption mass spectrometry, or matrix assisted laser desorption ionization (MALDI) mass spectrometry techniques. MALDI is preferable because of its higher sensitivity and greater tolerance to contaminating substances from 2-D gels (Janies *et al.*, 1993; Morz *et al.*, 1994; Puppin, Hojrup and Bleasby, 1993). Furthermore, recent modifications to sample preparation methods have largely solved early difficulties experienced with the calibration of MALDI spectra (Morz *et al.*, 1994; Vorm and Mann, 1994; Vorm, Roepstorff and Mann, 1994). The high sensitivity of mass spectrometry allows a small fraction of a digest of a 1 ug protein spot to be used for analysis, and analysis itself is complete in a few minutes.

A major challenge associated with peptide mass fingerprinting is data interpretation prior to computer matching against libraries of theoretical peptide digests. Spectra must be examined carefully to determine which peaks represent peptide masses of interest, as there are often enzyme autodigestion products and contaminating substances present (Henzel *et al.*, 1993; Morz *et al.*, 1994; Rasmussen *et al.*, 1994). Furthermore, if protein alkylation and reduction has not been undertaken prior to protein digestion, peptide sequence coverage may be poor (40% to 70%), with some masses present representing disulfide bonded peptides originally present in the protein (Morz *et al.*, 1994). For eukaryotes, a serious issue is the alteration of peptide masses by the presence of post-translational modifications (Table 6). The mass of the unmodified peptide alone can be very difficult to determine. Two artificial modifications introduced by electrophoresis, an acrylamide adduct to cysteine and the oxidation of methionine, are also known to alter peptide masses (le Maire *et al.*, 1993; Hess *et al.*, 1993).

Table 6: Masses of some common post-translational modifications. Peptides carrying post-translational modifications complicate data analysis for peptide mass fingerprinting protein identification. This is especially so for protein glycosylation, which involves many different combinations of the hexosamines, hexoses, deoxyhexoses, and sialic acid.

Post-translational modification	Mass change
Acetylation	
* Acrylamide adduct to cysteine	- 42.04
Carboxylation of Asp or Glu	- 71.00
Deamidation of Asn or Gln	- 22.01
Disulfide bond formation	- 0.98
Deoxyhexoses (Fuc)	- 2.02
Formylation	136.14
Hexosamines (GlcN, GalN)	- 28.01
Hexoses (Glc, Gal, Man)	- 161.16
Hydroxylation	- 162.14
N-acetylhexosamines (GlcNAc, GalNAc)	- 16.00
Oxidation of Met	- 203.19
Phosphorylation	- 16.00
Pyroglutamic acid formed from Gln	- 70.98
Sialic acid (NeuNAc)	- 17.03
Sulfation	- 291.26
	+ 80.08

Table modified from Finnigan LASERFAT application data sheet 5
Asterisk * shows modifications that can arise artificially from the 2-D electrophoresis process

A number of computer programs are available for matching peptide maps against databases (reviewed in Correll, 1994). Matching is usually undertaken in an interactive manner, whereby peaks of mass 500–3000 Da are selected and matched under various search parameters including MW of protein, mass accuracy of peptides, and number of missed enzyme cleavages allowed (Henzel *et al.*, 1993; Mortz *et al.*, 1994; Rasmussen *et al.*, 1994). The correct protein identity is the protein which has the most peptide masses in common with the unknown sample. Identities have been established with as few as three peptides, but unambiguous identification is thought to require a mass spectrometric map covering most peptides of the protein (Mortz *et al.*, 1994; Yates *et al.*, 1993). To date, peptide mass fingerprinting of proteins has been undertaken from the human myocardial protein and keratinocyte maps, from an *E. coli* 2-D gel, and from reference maps of *Spiroplasma melliferum* and *Mycoplasma genitalium* (Sutton *et al.*, 1995; Rasmussen *et al.*, 1994; Henzel *et al.*, 1993; Cordwell *et al.*, 1995; Wasinger *et al.*, 1995), although the technique is most powerful when used in combination with another protein identification technique (Rasmussen *et al.*, 1994; Cordwell *et al.*, 1995).

MASS SPECTROMETRY SEQUENCE TAGGING

An extension of peptide mass fingerprinting has recently been described, called peptide sequence tagging (Mann and Wilm, 1994; Mann, 1995). This uses tandem mass spectrometry (MS/MS) to initially determine the mass of peptides, then subject them to fragmentation by collision with a gas, and finally determine the mass of fragments. The resulting spectra gives information about a peptide's amino acid sequence. The fragmentation masses of peptides can rarely be used to assign a complete sequence, but it usually allows a short 'sequence tag' of 2 or 3 amino acids to be determined. This sequence tag and the original peptide mass is matched by computer against a database, providing a likely identity of the peptide and the protein it came from. The major drawback for this technique as a mass screening tool is the complexity of the mass data generated and the high level of expertise required for its interpretation. Nevertheless, it represents a useful new protein identification method which greatly increases the power of peptide mass fingerprinting protein identification.

Cross-species protein identification

Protein sequence databases continue to grow at a rapid rate, yet it is not widely appreciated that close to 90% of all information contained in current protein databases comes from only 10 species (A. Bairoch, Pers. Comm.). Fortunately, this information can be used to study proteomes of organisms that are poorly defined at the molecular level, via 2-D electrophoresis and 'cross-species' protein identification (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995). This approach allows proteins from reference maps of many different species to be identified without the need for the corresponding genes to be cloned and sequenced. This is particularly true for 'housekeeping' proteins, such as enzymes involved in glycolysis, DNA manipulation and protein manufacture, which are highly conserved across species boundaries. Proteins that cannot be identified across species boundaries can then become the focus of further protein characterisation and DNA sequencing efforts.

Figure
and try
to com
identifi
and Hi
all ent
protein
match
apulip

A)

Protein APAL_HUMAN

Axx: 8.4 Cxx: 19.3 Ser: 6.3 His: 1.3
 Gly: 4.2 Thr: 4.3 Ala: 8.0 Pro: 4.2
 Tyr: 2.9 Isg: 6.7 Val: 5.5 Met: 1.1
 Cys: 0.0 Leu: 25.5 Phe: 2.5 Lys: 8.8

pI Range: no range specified
 MW Range: no range specified

The closest SWISS-PROT entries are:

Rank	Score	Protein	(pI)	MW	Description
1	6	APAL_HUMAN	5.27	28078	APOLIPOPROTEIN A-I.
2	4	APAL_MACACA	5.43	28005	APOLIPOPROTEIN A-I.
3	10	APAL_RABBIT	5.15	27836	APOLIPOPROTEIN A-I.
4	14	APAL_BOTTIN	5.36	27549	APOLIPOPROTEIN A-I.
5	14	APAL_CANFIA	5.10	27467	APOLIPOPROTEIN A-I.
6	18	APAL_MOUSE	5.42	27322	APOLIPOPROTEIN A-I.
7	24	APAL_PIG	5.19	27598	APOLIPOPROTEIN A-I.
8	27	APAL_CHICK	5.26	27966	APOLIPOPROTEIN A-I.
9	37	DYNA_CHICK	5.44	117742	DYNACTIN. 117 KD ISOFORM.
10	39	APAC_HUMAN	5.18	43374	APOLIPOPROTEIN A-IV.

B)

Reagent: Trypsin MW filter: 10%

Scan using fragment mws cf:

1953	1933	1731	1613	1401	1387
1301	1283	1252	1235	1231	1215
1031	896	673	831	813	781
732	704				

No. of database entries scanned = 72018

1	. APAL_HUMAN	APOLIPOPROTEIN A-I (APO-AI). - HOMO SAPIENS
2	. APAL_MACFA	APOLIPOPROTEIN A-I (APO-AI). - MACACA FASCICULARIS
3	. APAL_PAPHA	APOLIPOPROTEIN A-I (APO-AI). - PAPIO HAMADRYAS
4	. B61845	cff B - Treponema denticola
5	. APA1_CANFA	APOLIPOPROTEIN A-I (APO-AI). - CANIS FAMILIARIS (DOG).
6	. S30947	hypothetical protein ; - Aizotobacter vinelandii
7	. HS2C_PEA	CHLOROPLAST HEAT SHOCK PROTEIN PRECURSOR. - PISUM SATIVU
8	. S20724	Tropomyosin - African clawed frog
9	. HIVV1354	HIVV1354 premature term. at 793 - Human immunodeficiency
10	. TRAJ_ECOLI	TRAJ PROTEIN. - ESCHERICHIA COLI.

Figure 6. Theoretical cross-species matching of human apolipoprotein A-I by amino acid composition and tryptic peptides. When an unknown protein is analyzed, best ranking proteins from both techniques can be compared. If the same protein type is observed in both lists, there is high confidence in this being the identity of the unknown molecule (Cordwell *et al.*, 1995; (A) Output of ExPASy server (Appel, Barich and Hochstrasser, 1994) where the true amino acid composition of apolipoprotein A-I was matched against all entries in the SWISS-PROT database, without pI or MW windows. Seven of the top 10 matching proteins were apolipoprotein A-I from different species. (B) Output of MOWSE peptide mass fingerprinting program (Pappin, Hojrup and Bleasby, 1993) where true tryptic peptides of human apolipoprotein A-I were matched against the OWL database, using MW window of 10%. Four of the top ten matching proteins were apolipoprotein A-I from different species.

Rapid cross-species identification of proteins from 2-D reference maps can be undertaken with amino acid composition or peptide mass fingerprinting methods (Figure 6), but these techniques alone may not identify proteins unambiguously when phylogenetic cross-species distances are great or analysis data is of poor quality (Yiye *et al.*, 1993; Shaw, 1993; Cordwell *et al.*, 1995). However, very high confidence in protein identities can be achieved when lists of best-matching proteins generated by both techniques are compared (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995). The correct identification is found when the same protein is ranked highly in lists of best matches generated by both techniques. This method has allowed approximately 120 proteins from the reference map of the mollicute *Spiroplasma melliferum*, representing approximately one quarter of the proteome, to be confidently identified by reference to protein information from other species (S. Cordwell, Personal Communication). When cross-species protein identification is to be undertaken, it should be noted that the molecular weight of a protein type across species is usually highly conserved, but that protein pI can vary by more than 2 units (Cordwell *et al.*, 1995). Accurate molecular weight determination by direct mass spectrometry of proteins blotted to PVDF (Eckerskorn *et al.*, 1992) should therefore be a useful additional parameter for cross-species protein identification.

CHARACTERISATION OF POST-TRANSLATIONAL MODIFICATIONS

Many proteins are modified after translation. Such post-translational modifications, including glycosylation, phosphorylation, and sulfation (see Table 6), are usually necessary for protein function or stability. Some abnormal modifications are associated with disease (Duthel and Revol, 1993; Ghosh *et al.*, 1993; Yamashita *et al.*, 1993). In proteome studies, post-translational modifications can be examined on all proteins present, or on individual spots. Studies on all proteins provide an indication of which proteins may carry a certain type of modification. For example, 2-D gel analysis of cell cultures grown in the presence of [³H]mannose or [³²P]phosphate gives an indication of which proteins carry glycans containing mannose, and which proteins are phosphorylated (Guttes and Franza, 1989). Lectin binding studies of 2-D gels blotted to PVDF or nitrocellulose provide information on the saccharides, if any, that are carried by proteins present (Gravel *et al.*, 1994).

When individual proteins of interest carrying post-translational modifications have been found, micropreparative 2-D electrophoresis can be used to purify them in microgram quantities (Hanash *et al.*, 1991; Bjellqvist *et al.*, 1993b). If protein isoforms of similar MW and pI are to be studied, focusing with narrow range pI gradients (1 pH unit) can provide greater separation and resolution. After electrophoresis, the type and degree of protein phosphorylation can be investigated (Murthy and Iqbal, 1991; Gold *et al.*, 1994), monosaccharide composition can be determined (Weitzhandler *et al.*, 1993; Packer *et al.*, 1995), and the structure and exact site of glycoamino acids can be investigated by either Edman degradation based techniques or by mass spectrometry (Pisano *et al.*, 1993; Hubeny *et al.*, 1993; Carr, Huddleston and Bean, 1993). With further development of rapid techniques, investigation of phosphorylation and monosaccharides by chromatographic or mass spectrometric means is likely to become a routine step in the characterisation of post-translational modifications of proteins from reference maps.

Table *
PROT
referen
from B
1946.

Species

Mycop.
Escheri
Sacchari
Diction
Arribalz
Cuenca
Hemps.

The under
becau
hundr
estim:
to less
protei
electri-
ism c
acceler
are ur
post-l
diff n
usefu:

The status of proteome projects

Many technical aspects of proteome research have already been discussed in this review, but an overview of the status of proteome projects has not yet been presented. Advances in proteome projects will initially rely on progress in genome sequencing initiatives, to enable an identity, amino acid sequence, or function to be assigned to each protein spot. Table 7 shows genome size, proteome size, and the number of proteins already defined for a number of model organisms. This indicates that whilst genome sequencing programs for *E. coli* and *S. cerevisiae* are advanced, the massive size of some other genomes (and especially the human genome) means that their complete nucleotide sequences are unlikely to be available for many years. Because of this, 2-D reference maps and proteome projects of single cell organisms like *Mycoplasma* sp., *E. coli* and *S. cerevisiae* will be the most detailed (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995; Vanbogelen *et al.*, 1992; Gartels *et al.*, 1994), and complete maps of other organisms will take longer to construct. However, the use of cross-species protein identification techniques will allow proteomes of many prokaryotes and simple eukaryotes to be partially defined in reference to *E. coli* and *S. cerevisiae*.

Table 7: Estimated genome size, estimated proteome size, number of protein sequences in SWISS-PROT Release 31 (March, 1995), and approximate number of proteins of known identity on 2-D reference maps for some model organisms. Genome size data from Smith (1993), and total protein data from Bird (1995). Genome sequencing projects of *E. coli* and *S. cerevisiae* will probably be complete in 1996.

Species Name	Haploid genome size (million bp)	Estimated proteome size (total proteins)	Protein entries in SWISS PROT	Proteins annotated on 2-D Maps
<i>Mycoplasma</i> species	0.6–0.8	4(XX)–600	1(X)	> 1(X)
<i>Escherichia coli</i>	4.8	4000	3170	> 3(X)
<i>Saccharomyces cerevisiae</i>	13.5	6000	3160	> 1(X)
<i>Drosophila melanogaster</i>	70	12500	214	–
<i>Acanthocystis turfacea</i>	70	14(XX)	270	–
<i>Cucurbita sativa</i>	80	17000	702	–
<i>Human sapiens</i>	2900	60000–80000	3326	> 1(XXX)

The study of vertebrate proteomes and vertebrate development is a phenomenal undertaking in comparison to the investigation of single cell organisms. This is because vast numbers of proteins are developmentally expressed, each body tissue has hundreds of unique proteins, and there are numerous tissue types. However, it is estimated that at least 35% of proteins in vertebrate cells will be conserved from tissue to tissue, constituting the 'housekeeping' proteins (Bird, 1995), with the remainder of proteins constituting a set that are specific to a cell type. Providing that standardised electrophoretic conditions are used, reference maps from many tissues of one organism can be superimposed in gel databases (e.g. Hochstrasser *et al.*, 1992). This accelerates the definition of the 'housekeeping' proteins, as well as sets of proteins that are unique to different tissue types. Such studies may, however, be complicated by post-translational modifications, which can differ on the same gene product in different tissues. Proteins that remain unknown after identification procedures will be useful in providing focus for nucleic acid sequencing initiatives.

FUTURE DIRECTIONS OF PROTEOME PROJECTS

This review has described recent advances in the area of proteome research. It has illustrated how new developments of older techniques (2-D electrophoresis and amino acid analysis) as well as the applications of new technology (mass spectrometry) have greatly widened the choice of tools the biologist and protein chemist has for the separation, identification and analysis of complex mixtures of proteins. This has made possible the establishment of detailed reference maps for organisms, which are becoming the method of choice for the definition of tissues or whole cells, and the investigation of gene expression therein.

Proteome projects are already impacting on the dogma of molecular biology that DNA sequence constitutes the definition of an organism. For example, the proteomes of different tissues of a single organism are often significantly different. Similarly, cross-species identification of proteins (for example the identification of proteins from *Candida albicans* by comparison with *S. cerevisiae*) can open up studies on organisms that are poorly molecularly defined. As cross-species identification can proceed at a pace orders of magnitude faster than a genome project in terms of defining the gene and protein complement of organisms, the need for the DNA sequencing of genomes will be avoided, and emphasis placed on those found to be novel.

Just as genome sequencing is not an end in itself, neither is an annotated 2-D protein reference map of an organism, nor indeed the identification of proteins in a proteome. So whilst an immediate aim of proteome projects is to screen proteins in reference maps, this will lead to expression studies and characterisation of post-translational modifications. The challenge that then needs to be addressed is the investigation of structure and function of proteins in a proteome. The magnitude of this is illustrated by the fact that over half the open reading frames identified in *S. cerevisiae* chromosome III were initially of no known function (Oliver *et al.*, 1992). Structural and functional studies will be an undertaking just as formidable as genome studies are now and proteome projects are becoming, but will lead to an unimaginably detailed understanding of how living organisms are constructed and how they operate.

Acknowledgements

MRW is the recipient of an Australian Postgraduate Research Award. AAG, MRW, IHS and KLW acknowledge assistance for proteome projects through Macquarie University Research Grants, the Australian Research Council, the Australian National Health and Medical Research Council, Beckman Instruments and GBC Scientific Equipment. DH acknowledges the financial support of a Monash Foundation Grant and the Swiss National Fund for Scientific Research (Grant # 31-53658.92). We thank colleagues who supplied work that was 'In Press' during the writing of this review.

References

- ANDERSON, N.L., HOFMANN, J.P., GEMMELL, A. AND TAYLOR, J. (1984). Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis. *Clinical Chemistry*, 30, 2031-2036.

- ANDERSON, N.L. AND ANDERSON, N.G. (1991). A two-dimensional gel database of human plasma proteins. *Electrophoresis*, 12, 883-906.
- ANDERSON, N.L., ESQUER-BLASCO, R., HOFMANN, J.P. AND ANDERSON, N.G. (1991). A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effect studies. *Electrophoresis*, 12, 907-930.
- ANDERSON, N.L., COPPLE, D.C., BENDELE, R.A., PROBST, G.S., RICHARDSON, F.C. (1992). Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. *Environmental and Applied Toxicology*, 18, 570-580.
- APPEL, R.D., BAIROCH, A. AND HOCHSTRASSER, D.F. (1992). A new generation of information retrieval tools for biologists: the example of the EXPASY WWW server. *Trends in Biochemical Sciences*, 19, 256-260.
- APPEL, R.D., HOCHSTRASSER, D.F., FLUNK, M., VARGAS, J.R., PELIGRINI, C., MULLER, A.F. AND SCHERRER, J.R. (1991). The MELANIE project: from a biopsy to automatic protein map interpretation by computer. *Electrophoresis*, 12, 722-735.
- APPEL, R.D., SANCHEZ, J.-C., BAIROCH, A., GOLAZ, O., MILI, M., VARGAS, J.R. AND HOCHSTRASSER, D.F. (1993). SWISS-2DPAGE: a database of two-dimensional gel electrophoresis images. *Electrophoresis*, 14, 1323-1328.
- APPEL, R.D., SANCHEZ, J.-C., BAIROCH, A., GOLAZ, O., RAVIER, F., PASQUALI, C., HUGHES, G. AND HOCHSTRASSER, D.F. (1994). The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis. *Nucleic Acids Research*, 22, 3581-3582.
- BAIROCH, A. AND BOECKMANN, B. (1994). The SWISS-PROT protein sequence databank: current status. *Nucleic Acids Research*, 22, 3578-3580.
- BAKER, C.S., CORBETT, J.M., MAY, A.J., YACOUB, M.H. AND DUNN, M.J. (1992). A human myocardial two-dimensional electrophoresis database: protein characterisation by microsequencing and immunoblotting. *Electrophoresis*, 13, 723-726.
- BARTLET-JONES, M., JEFFERY, W.A., HANSEN, H.F. AND PAPPIN, D.J.C. (1994). Peptide ladder sequencing by mass spectrometry using a novel volatile degradation reagent. *Rapid Communications in Mass Spectrometry*, 8, 737-742.
- BAUER, D., MULLER, H., REICH, J., RIEDEL, H., AHRENKIEL, V., WARTHOE, P. AND STRASS, M. (1993). Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Research*, 21, 4273-4280.
- BERNERS-LEE, T.J., CAILIAU, R., GROFF, J.F. AND POLLERNAN, B. (1992). *Electronic Networks: Research, Applications, and Policy*, 2, 52-58.
- BIRD, A.P. (1995). Gene number, noise reduction and biological complexity. *Trends in Genetics*, 11, 9-100.
- BJELLOVIST, B., EK, K., RIGHETTI, P.G., GIANAZZA, E., GORG, A., WESTERNIEIER, R. AND POSTEL, W. (1982). Isoelectric focusing in immobilized pH gradients principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*, 6, 317-339.
- BJELLOVIST, B., PASQUALI, C., RAVIER, F., SANCHEZ, J.-C. AND HOCHSTRASSER, D.F. (1993a). A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis*, 14, 1357-1365.
- BJELLOVIST, B., SANCHEZ, J.-C., PASQUALI, C., RAVIER, F., PAQUET, N., FRUTIGER, S., HUGHES, G.J. AND HOCHSTRASSER, D.F. (1993b). Micropreparative 2-D electrophoresis allowing the separation of milligram amounts of proteins. *Electrophoresis*, 14, 1375-1378.
- BJELLOVIST, B., HUGHES, G., PASQUALI, C., PAQUET, N., RAVIER, F., SANCHEZ, J.-C., FRUTIGER, S. AND HOCHSTRASSER, D. (1993c). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14, 1023-1031.
- BONNER, W.M. AND LASKEY, R.A. (1974). A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. *European Journal of Biochemistry*, 46, 83-88.
- BOUTELL, T., GARRELS, J.I., FRANZA, B.R., MONARDO, P.J. AND LATTER, G.J. (1994). REF52 on global gel navigator: an internet-accessible two-dimensional gel electrophoresis database. *Electrophoresis*, 15, 1487-1490.
- BREWER, J., GRUND, E., HAGERLID, P., OLSSON, I. AND LIZANA, J. (1986). In *Electrophoresis '86* (M.J. Dunn, Ed.), pp. 226-229. VCH, Weinheim.

- CARR S.A., HUDDLESTON, M.J. AND LEAN, M.F. (1993). Selective identification and differentiation of N- and O-linked oligosaccharides in glycoproteins by liquid chromatography-mass spectrometry. *Protein Science*, 2, 183-196.
- CARTERILL, K.P., MACLENNAN, D.H. AND JORGENSEN, A.O. (1983). Staining of the Cu-binding proteins, calnexin, calreticulin, calmodulin, troponin C, and S-100, with the cationic dye 'Stain-all'. *Journal of Biochemical Chemistry*, 258, 11267-11273.
- CELIS, J.E., CRUGER, D., KIEL, J., DEJGAARD, K., LARIDSEN, J.B., RATZ, G.P., BASSE, B., CELIS, A., RASMUSSEN, H.H., BALW, G. AND VANDEKERKHOVE, J. (1990a). A two-dimensional gel protein database of uncultured total normal human epidermal keratinocytes: identification of proteins strongly up-regulated in psoriatic epidermis. *Electrophoresis*, 11, 242-254.
- CELIS, J.E., GESSER, B., RASMUSSEN, H.H., MADSEN, P., LEFFERS, H., DEJGAARD, K., HONORE, B., OLSEN, E., RATZ, G., LARIDSEN, J.B., BASSE, B., MORITZEN, S., HELLERUP, M., ANDERSEN, A., WALBURN, E., CELIS, A., BALW, G., PLYPE, M., VAN DAMME, J. AND VANDEKERKHOVE, J. (1990b). Comprehensive two-dimensional gel protein database offers a global approach to the analysis of human cells: the transformed amnion cells (TANIA) master database and its link to genome DNA sequence data. *Electrophoresis*, 11, 898-1071.
- CELIS, J.E., RASMUSSEN, H.H., OLSEN, E., MADSEN, P., LEFFERS, H., HONORE, B., DEJGAARD, K., GROMOV, P., HOFFMANN, H.J., NIELSEN, M., VASSILEV, A., VINTERMYR, O., HAO, J., CELIS, A., BASSE, B., LAURIDSEN, J., RATZ, G.P., ANDERSEN, A.H., WALBURN, E., KJAERGAARD, I., PLYPE, M., VAN DAMME, J., AND VANDEKERKHOVE, J. (1993). The human keratinocyte two-dimensional database: update 1993. *Electrophoresis*, 14, 1091-1195.
- CELIS, J.E. AND OLSEN, E. (1994). A qualitative and quantitative protein database approach identified individual and groups of functionally related proteins that are differentially regulated in simian virus 40 (SV40) transformed human keratinocytes: an overview of the functional changes associated with the transformed phenotype. *Electrophoresis*, 15, 309-324.
- CORBETT, J.M., DUNN, M.J., POSCH, A. AND GORG, A. (1994a). Positional reproducibility of protein spots in two-dimensional polyacrylamide electrophoresis using immobilized pH gradient isoelectric focusing in the first dimension - an interlaboratory comparison. *Electrophoresis*, 15, 1205-1211.
- CORBETT, J.M., WHEELER, C.H., BAKER, C.S., YACOUB, M.H. AND DUNN, M.J. (1994b). The human myocardial two-dimensional gel protein database: update 1994. *Electrophoresis*, 15, 1459-1465.
- CORDWELL, S., WILKINS, M.R., CERPA-POLAK, A., GOOLEY, A.A., DUNCAN, M., WILLIAMS, K.L. AND HUMPHREY-SMITH, I. (1995). Cross-species identification of proteins separated by two-dimensional electrophoresis using MALDI-TOF and amino acid composition. *Electrophoresis*, 15, 238-243.
- COTTRELL, J.S. (1994). Protein identification by peptide mass fingerprinting. *Peptide Research*, 7, 115-124.
- CRIMMINS, D.L., MCCOURT, D.W., THOMA, R.S., SCOTT, M.G., MACKE, K. AND SCHWARTZ, B.D. (1990). *In situ* chemical cleavage of proteins immobilized to glass-fiber and polyvinylidenefluoride membranes: cleavage at tryptophan residues with 2-(2-nitrophenylsulfonyl)-3-methyl-3-bromoindolenine to obtain internal amino acid sequence. *Analytical Biochemistry*, 187, 27-38.
- DUTHEL, S. AND REVOL, A. (1993). Glycan microheterogeneity of alpha 1-antitrypsin in serum and meconium from normal and cystic fibrosis patients by crossed immunoelectrophoresis with different lectins (Con A, LCA, WGA). *Clinical and Chemical Acta*, 215, 173-187.
- ECKERSKORN, C., JUNGBLIT, P., MEWES, W., KLOSE, J. AND LOTTSPEICH, F. (1988). Identification of mouse brain proteins after two-dimensional electrophoresis and electroblotting by microsequence analysis and amino acid composition. *Electrophoresis*, 9, 830-838.
- ECKERSKORN, C., STRUPAT, K., KARAS, M., HILLENKAMP, F. AND LOTTSPEICH, F. (1992).

- Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis*, 13, 664-665.
- ECKERSKORN, C. AND LOTTspeICH, F. (1993). Structural characterisation of blotting membranes and the influence of membrane parameters for electroblotting and subsequent amino acid sequence analysis of proteins. *Electrophoresis*, 14, 831-836.
- ERIK, K., BJELLOVIST, B.J. AND RIGHETTI, P.G. (1983). Preparative isoelectric focusing in immobilized pH gradients. I. General principle and methodology. *Journal of Biochemical and Biophysical Methods*, 8, 135-155.
- FEY, S.J., CARLSEN, J., MOSE LARSEN, P., JENSEN, U.A., KJELDSEN, K. AND HALVØS, S. (1992). Two-dimensional gel electrophoresis-a tool for molecular cardiology. Proceedings of the International Society for Heart Research 'XV European Section Meeting', pp 9-16.
- FREY, J.R., KUHN, L., KETTMAN, J.R. AND LEFKOVITS, I. (1994). The amino acid composition of 350 lymphocyte proteins. *Molecular Immunology*, 31, 1219-1231.
- GARRELS, J.I. (1989). The QUEST system for quantitative analysis of two-dimensional gels. *Journal of Biological Chemistry*, 264, 5269-5282.
- GARRELS, J.I. AND FRANZA, B.R. (1989). The REF52 protein database. *Journal of Biological Chemistry*, 264, 5283-5298.
- GARRELS, J.I., FRANZA, B.R., CHANG, C. AND LATTER, G. (1990). Quantitative exploration of the REF52 protein database: cluster analysis reveals the major protein expression profiles in response to growth regulation, serum stimulation, and viral transformation. *Electrophoresis*, 11, 1114-1130.
- GARRELS, J.I., FUTCHER, B., KOBAYASHI, R., LATTER, I., SCHWENDER, B., VOLPE, T., WARNER, J.R. AND MCLAUGHLIN, C.S. (1994). Protein identification for a *Saccharomyces cerevisiae* protein database. *Electrophoresis*, 15, 1466-1486.
- GELFI, C., BOSSI, M.L., BJELLOVIST, B. AND RIGHETTI, P.G. (1987). Isoelectric focusing in immobilized pH gradients in the pH 10-11 range. *Journal of Biochemical and Biophysical Research Methods*, 15, 41-48.
- GHARAHDAGHI, F., ATHERTON, D., DEMOTT, M. AND MISCHE, S.M. (1992). Amino acid analysis of PVDF-bound proteins. In *Techniques in Protein Chemistry III* (R.H. Aegeletti, Ed.), pp 249-260. Academic Press, San Diego.
- GHOSH, P., OXOH, C., LIL, Q.H. AND LAKSHMIN, M.R. (1993). Effects of chronic ethanol on enzymes regulating sialylation and desialylation of transferrin in rats. *Alcoholism: Clinical and Experimental Research*, 17, 576-579.
- GIOVINETTI, C.S., GENMELL, M.A., TOLLAKSEN, S.L. AND TAYLOR, J. (1991). Quantitation of human leukocyte proteins after silver staining: a study with two-dimensional electrophoresis. *Electrophoresis*, 12, 536-543.
- GIOVINETTI, C.S., TAYLOR, J. AND TOLLAKSEN, S.L. (1992). Mouse liver protein database: a catalog of proteins detected by two-dimensional gel electrophoresis. *Electrophoresis*, 13, 970-991.
- GOLAZ, O., HUGHES, G.J., FRITZIGER, S., PAQUET, N., BAIROCH, A., PASOLALI, C., SANCHEZ, J.C., TISSOT, J.D., APPEL, R.D., WALZER, C., BALANT, L. AND HOCHSTRASSER, D.F. (1993). Plasma and red blood cell protein maps: update 1993. *Electrophoresis*, 14, 1223-1231.
- GOLD, M.R., YUNGWIRTH, T., SUTHERLAND, C.L., INGHAM, R.J., VIANONI, D., CHIU, R., VAN OOSTVEEN, L., MORRISON, H.D. AND AEBERSOLD, R. (1993). Purification and identification of tyrosine-phosphorylated proteins from lymphocytes stimulated through the antigen receptor. *Electrophoresis*, 15, 441-453.
- GOLDHBERG, H.A., DOMENICUCCI, C., PRINGLE, G.A. AND SODEK, J. (1988). Mineral-binding proteoglycans of fetal porcine calvarial bone. *Journal of Biological Chemistry*, 263, 12092-12101.
- GOOLEY, A.A., MARSHALEK, R. AND WILLIAMS, K.L. (1992). Size polymorphisms due to changes in the number of O-glycosylated tandem repeats in the *Dicystostelium discoideum* glycoprotein P_gA. *Genetics*, 130, 749-756.
- GORG, A., POSTEL, W. AND GUNTHER, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, 9, 531-546.
- GORG, A., POSTEL, W., GUNTHER, S., WESER, J., STRAHLER, J.R., HANASII, S.M., SOYERLOT, L.

- AND KLICK, R. (1988). Approach to stationary two-dimensional pattern: influence of focusing time and imidobilized carrier ampholyte concentrations. *Electrophoresis*, 9, 37-46.
- GRAVEL, P., GOLAZ, O., WALTZER, C., HOCHSTRASSER, D.F., TURLER, H., AND BALANT, L.P. (1994). Analysis of glycoproteins separated by two-dimensional gel electrophoresis using lectin blotting revealed by chemiluminescence. *Analytical Biochemistry*, 221, 66-71.
- GUTHIER, S., POSTEL, W., WILRING, H. AND GORG, A. (1988). Acid phosphatase typing for breeding nematode-resistant tomatoes by isoelectric focusing with an ultranarrow immobilized pH gradient. *Electrophoresis*, 9, 616-620.
- HANASH, S.M., STRAHLER, J.R., NEEL, J.V., HAILAT, N., MELHEMI, R., KEIM, D., ZHI, X.X., WAGNER, D., GAGE, D.A. AND WATSON, J.T. (1991). Highly resolving two-dimensional gels for protein sequencing. *Proceedings of the National Academy of Sciences U.S.A.*, 88, 5709-5713.
- HARRINGTON, M.G., COFFMAN, J.A., CALZONE, F.J., HOOD, L.E., BRITTON, R.J. AND DAVIDSON, E.H. (1992). Complexity of sea urchin embryo nuclear proteins that contain basic domains. *Proceedings of the National Academy of Sciences U.S.A.*, 89, 6252-6256.
- HARRINGTON, M.G., LEE, K.H., YU, M., ZEWERT, T., BAILEY, J.E. AND HOOD, L.E. (1993). Mechanical precision in two-dimensional electrophoresis can improve spot positional reproducibility. *Applied and Theoretical Electrophoresis*, 3, 347-353.
- HELLMAN, U., WERNSTEDT, C., GOMEZ, J. AND HELDIN, C.-H. (1995). Improvement of an in-gel digestion for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical Biochemistry*, 224, 451-455.
- HENZEL, W.J., BILLECI, T.M., STULTS, J.T., WONG, S.C., GRIMLEY, C. AND WATANABE, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences U.S.A.*, 90, 5011-5015.
- HESS, D., COVEY, T.C., WINZ, R., BROWNSEY, R.W. AND AEBERSOLD, R. (1993). Analytical and micropreparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis. *Protein Science*, 2, 1342-1351.
- HOBOHM, U., HOLTHAEVE, T. AND SANDER, C. (1994). Amino acid analysis and protein database compositional search as a rapid and inexpensive method to identify proteins. *Analytical Biochemistry*, 222, 202-209.
- HOCHSTRASSER, D.F. AND MERRIL, C.R. (1988). 'Catalysts' for polyacrylamide gel polymerization and detection of proteins by silver staining. *Applied and Theoretical Electrophoresis*, 1, 35-40.
- HOCHSTRASSER, D.F., PATCHORNICK, A. AND MERRIL, C.R. (1988). Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Analytical Biochemistry*, 173, 412-423.
- HOCHSTRASSER, A.C., JAMES, R.W., PONETTA, D. AND HOCHSTRASSER, D.F. (1991a). Preparation, livezone electrofocusing and high resolution two-dimensional electrophoresis for concentration and purification of proteins. *Applied and Theoretical Electrophoresis*, 1, 333-337.
- HOCHSTRASSER, D.F., APPEL, R.D., VARGAS, R., PERKIER, R., VILROUD, J.F., RAVIER, F., PASQUALI, C., FUNK, M., PELLIGRINI, C., MULLER, A.F. AND SCHERRER, J.R. (1991b). A clinical molecular scanner: the Melanie project. *Medical Computing*, 8, 85-91.
- HOCHSTRASSER, D.F., FRUTIGER, S., PAQUET, N., BAIROCH, A., RAVIER, F., PASQUALI, C., SANCHEZ, J.-C., TISSOT, J.-D., BJELLOVIST, B., VARGAS, R., APPEL, R.D. AND HUGHES, G.J. (1992). Human liver protein map: a reference database established by microsequencing and gel comparison. *Electrophoresis*, 13, 992-1001.
- HOLT, T.G., CHANG, C., LAURENT-WINTER, C., MURAKAMI, T., DAVIES, J.E. AND THOMPSON, C.J. (1992). Global changes in gene expression related to antibiotic synthesis in *Streptomyces hygroscopicus*. *Molecular Microbiology*, 6, 969-980.
- HONORE, B., LEFFERS, H., MADSEN, P. AND CELIS, J.E. (1993). Interferon-gamma up-regulates a unique set of proteins in human keratinocytes. Molecular cloning and expression of the cDNA encoding the RGD-sequence containing protein IGUP 1-5111. *European Journal of Biochemistry*, 218, 421-430.
- HUBERTY, M.C., VATH, J.E., YU, M. AND MARTIN, S.A. (1993). Site-specific carbohydrate

- identification in recombinant proteins using MALDI-TOF MS. *Analytical Chemistry*, 65, 2791-2800.
- HUGHES, G.J., FRITZGER, S., PAOLIET, N., PASOLALI, C., SANCHEZ, J.-C., TISSOT, J.D., BAIROCH, A., APPEL, R.D. AND HOCHSTRASSER, D.F. (1993). Human liver protein map update 1993. *Electrophoresis*, 14, 1216-1222.
- HUGHES, J.H., MACK, K. AND HAMPIARIAN, V.V. (1988). India ink staining of proteins on nylon and hydrophobic membranes. *Analytical Biochemistry*, 173, 18-25.
- JANIES, P., QLADROVI, M., CARAFOLI, E. AND GONNET, G. (1993). Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications*, 195, 58-64.
- JI, H., WHITEHEAD, R.H., REID, G.E., MORITZ, F.L., WARD, L.D. AND SIMPSON, R.J. (1993). Two-dimensional electrophoretic analysis of proteins expressed by normal and cancerous human crypts: application of mass spectrometry to peptide-mass fingerprinting. *Electrophoresis*, 15, 391-395.
- JOHNSON, R.S. AND WALSH, K.A. (1992). Sequence analysis of peptide mixtures by automated integration of Edman and mass spectrometric data. *Protein Science*, 1, 1083-1091.
- JOHNSTON, R.F., PICKETT, S.C. AND BARKER, D.L. (1990). Autoradiography using storage phosphor technology. *Electrophoresis*, 11, 355-360.
- JUNGBLUT, P., DZIONARA, M., KLOSE, J. AND WITTMANN-LEIBOLD, B. (1992). Identification of tissue proteins by amino acid analysis after purification by two-dimensional electrophoresis. *Journal of Protein Chemistry*, 11, 603-612.
- JUNGBLUT, P., OTTO, A., ZEINDL-EBERHART, E., PLEIBNER, K.-P., KNECHT, M., REGITZ-ZAGROSEK, V., FLECK, E. AND WITTMANN-LEIBOLD, B. (1994). Protein composition of the human heart: the construction of a myocardial two-dimensional electrophoresis database. *Electrophoresis*, 15, 685-707.
- KOHARA, Y., AKIYAMA, K. AND ISONO, K. (1987). The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, 50, 495-508.
- KLOSE, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues. A novel approach to testing for individual point mutations in mammals. *Human Genetik*, 26, 231-243.
- LATHAM, K.E., GARRELS, J.I., CHANG, C. AND SOLTER, D. (1991). Quantitative analysis of protein synthesis in mouse embryos I: extensive re-programming at the one- and two-cell stages. *Development*, 2, 921-932.
- LATHAM, K.E., GARRELS, J.I., CHANG, C. AND SOLTER, D. (1992). Analysis of embryonic mouse development: construction of a high-resolution, two-dimensional gel protein database. *Applied and Theoretical Electrophoresis*, 2, 163-170.
- LATHAM, K.E., GARRELS, J.I. AND SOLTER, D. (1993). Two-dimensional analysis of protein synthesis. *Methods in Enzymology*, 255, 473-489.
- LE MAIRE, M., DESCHAUMPS, S., MOLLER, J.V., LE CAER, J.P. AND ROSSIER, J. (1993). Electrospray ionization mass spectrometry from sodium dodecyl sulfate-polyacrylamide gel electrophoresis: application to the topology of the sarcoplasmic reticulum Ca²⁺-ATPase. *Analytical Biochemistry*, 214, 50-57.
- LEVKIN, P.F. AND LESTER, E.P. (1989). Database and search techniques for two-dimensional gel protein data: a comparison of paradigms for exploratory data analysis and prospects for biological modelling. *Electrophoresis*, 10, 122-140.
- LEVKIN, P.F., WL, Y. AND UPTON, K. (1993). An efficient disk-based data structure for rapid searching of quantitative two-dimensional gel databases. *Electrophoresis*, 14, 1341-1350.
- LI, K.W., GERAERTS, W.P., VAN-ELK, R. AND KOOSE, J. (1989). Quantification of proteins in the subnanogram and nanogram range: comparison of the AutoDye, FerriDye, and india ink staining methods. *Analytical Biochemistry*, 182, 44-47.
- LIANG, P. AND PARDEE, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257, 967-971.
- MANN, M. (1995). Sequence database searching by mass spectrometric data. In *Microcharacterisation of Proteins* (R. Kellner, F. Lutzspeich, and H.E. Meyer, Eds), pp. 223-245. VCH, Weinheim.

- MANN, M., HOJERUP, P. AND ROEPSTORFF, P. (1993). Use of mass spectrometric molecular weight informatics to identify proteins in sequence databases. *Biological Mass Spectrometry*, 22, 338-345.
- MANN, M. AND WILKINS, M. (1994). Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry*, 66, 4390-4399.
- MATSUDAIRA, P. (1987). Sequence of picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry*, 262, 10035-10038.
- MONARDE, P.J., BOLTELL, T., GARRELS, J.I. AND LATTER, G.J. (1994). A distributed system for two-dimensional gel analysis. *Computer Applications in the Biosciences*, 10, 137-143.
- MORTZ, E., VOPAL, O., MANN, M. AND ROEPSTORFF, P. (1993). Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biological Mass Spectrometry*, 23, 249-261.
- MURTHY, L.R. AND JOBAL, K. (1991). Measurement of picomoles of phosphoamino acids by high performance liquid chromatography. *Analytical Biochemistry*, 193, 299-303.
- MYRICK, J.E., LEMKIN, P.F., ROBINSON, M.K. AND UPTON, K.M. (1993). Comparison of the BioImage Visage 2000 and the GELLAB-II two-dimensional electrophoresis image analysis systems. *Applied and Theoretical Electrophoresis*, 3, 335-346.
- NEIDHARDT, F.C., APPLEBY, D.B., SANKAR, P., HUTTON, M.E. AND PHILLIPS, T.A. (1989). Genomically linked cellular protein databases derived from two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis*, 10, 116-122.
- NIKODEM, V. AND FRESCO, J.R. (1979). Protein fingerprinting by SDS-gel electrophoresis after partial fragmentation with CNBr. *Analytical Biochemistry*, 97, 382-386.
- NOKIHARA, K., MORITA, N. AND KURIKI, T. (1992). Applications of an automated apparatus for two-dimensional electrophoresis. Model TEP-1, for microsequence analysis of proteins. *Electrophoresis*, 13, 701-707.
- O'FARRELL, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, 250, 4007-4021.
- O'FARRELL, P.Z., GOODMAN, H.M. AND O'FARRELL, P.H. (1977). High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell*, 12, 1133-1142.
- OLIVER *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature*, 357, 38-46.
- OLSEN, A.D. AND MILLER, M.J. (1988). Elsie 4: quantitative computer analysis of sets of two-dimensional gel electropherograms. *Analytical Biochemistry*, 169, 49-70.
- ORTIZ, M.L., CALERO, M., FERNANDEZ-PATRON, C., PATRON, C.F., CASTELLANOS, L. AND MENDEZ, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. *FEBS Letters*, 296, 300-304.
- OSTERGREW, K., ERIKSSON, G. AND BJELLOVIST, B. (1988). The influence of support material used on band sharpness in Immobiline gels. *Journal of Biochemical and Biophysical Methods*, 16, 165-170.
- OL, K., WILKINS, M.R., YAN, J.X., GOOLEY, A.A., FUNG, Y., SHELMACK, D. AND WILLIAMS, K.L. (1995). Improved high-performance liquid chromatography of amino acids derivatized with 9-fluorenylmethyl chloroformate. *Journal of Chromatography* (in press).
- PACKER, N., WILKINS, M.R., GOLAZ, O., LAWSON, M., GOOLEY, A.A., HOCHSTRASSER, D.F., REDMOND, J. AND WILLIAMS, K.L. (1995). Characterisation of human plasma glycoproteins separated by two-dimensional gel electrophoresis. *BioTechnology* (in press).
- PAPPIN, D.J.C., HOURUP, P. AND BLEASBY, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, 3, 327-332.
- PATTERSON, S.D. (1994). From electrophoretically separated protein to identification: strategies for sequence and mass analysis. *Analytical Biochemistry*, 221, 1-15.
- PATTERSON, S.D. AND LATTER, G.J. (1993). Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. *BioTechniques*, 15, 1076-1083.
- PISANO, A., REDMOND, J.W., WILLIAMS, K.L. AND GOOLEY, A.A. (1993). Glycosylation sites identified by solid-phase Edman degradation: O-linked glycosylation motifs on human glycophorin A. *Glycobiology*, 3, 429-435.
- RABILLON, D.T. (1992). A comparison between low background silver diamine and silver nitrate protein stains. *Electrophoresis*, 13, 429-439.

- RASMUSSEN, H.H., VAN DAVENNE, J., PUYPE, M., GESSER, B., CELIS, J.E. AND VANDENKERCKHOVE, J. (1992). Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes. *Electrophoresis*, 13, 960-969.
- RASMUSSEN, H.H., MORTZ, E., MANN, M., ROEPSTORFF, P. AND CELIS, J.E. (1994). Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass-spectrometric peptide mapping alone and in combination with microsequencing. *Electrophoresis*, 15, 406-416.
- RICHARDSON, F.C., HORN, D.M. AND ANDERSON, N.L. (1994). Dose-responses in rat hepatic protein modification and expression following exposure to the rat hepatocarcinogen methapyrilene. *Carcinogenesis*, 15, 325-329.
- RIGHETTI, P.G. (1990). Immobilized pH gradients: theory and methodology. In *Laboratory Techniques in Biochemistry and Molecular Biology* (R.H. Burdon and P.H. van Knippenberg, Eds), Elsevier, Amsterdam.
- RIGHETTI, P.G. AND DRYSDALE, J.W. (1973). *Annals of the New York Academy of Sciences*, 209, 163-186.
- RODRIGUEZ, L.V., GERNSTEN, D.M., RAMAGLI, L.S. AND JOHNSTON, D.A. (1993). Towards stoichiometric silver staining of proteins resolved in complex two-dimensional electrophoresis gels: real-time analysis of pattern development. *Electrophoresis*, 14, 628-637.
- ROSENFELD, J., CAPDEVIELLE, J., GUILLEMET, J.C. AND FERRARA, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Analytical Biochemistry*, 203, 173-179.
- SANCHEZ, J.C., RAVIER, F., PASQUALI, C., FRITZIGER, S., PAQUET, N., BJELLOVIST, B., HOCHSTRASSER, D.F. AND HUGHES, G.J. (1992). Improving the detection of proteins after transfer to polyvinylidene difluoride membranes. *Electrophoresis*, 13, 715-717.
- SANGER, F., COULSON, A.R., HONG, G.F., HILL, D.F. AND PETERSEN, G.B. (1982). Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, 162, 729-773.
- SCHEEL, G.J. (1975). Two-dimensional analysis of soluble proteins. *Biochemistry*, 250, 5375-5385.
- SHAW, G. (1993). Rapid identification of proteins. *Proceedings of the National Academy of Sciences USA*, 90, 5138-5142.
- SIBBALD, P.R., SOMMERFELDT, H. AND ARGOS, P. (1991). Identification of proteins in sequence databases from amino acid composition. *Analytical Biochemistry*, 198, 330-333.
- SIMPSON, R.J., TSUGITA, A., CELIS, J.E., GARRELS, J.I. AND MEWES, H.W. (1992). Workshop on two-dimensional gel protein databases. *Electrophoresis*, 13, 1053-1061.
- SINHA, P.K., KUTTER, E., STOFFLER, M.-M., GIANAZZA, E. AND RIGHETTI, P.G. (1990). Two-dimensional maps in very acidic immobilized pH gradients. *Journal of Biochemical and Biophysical Methods*, 20, 345-352.
- SMITH, D.W. (1994). Introduction. In *Biocomputing: Informatics and Genome Projects* (D.W. Smith, Ed.), pp1-12. Academic Press, San Diego.
- STRUPAT, K., KARAS, M., HILLENKAMP, F., ECKERSKORN, C. AND LOTTSPREICH, F. (1994). Matrix-assisted laser desorption ionization mass spectrometry of proteins electroblotted after polyacrylamide gel electrophoresis. *Analytical Chemistry*, 66, 464-470.
- SUTTON, C.W., PEMBERTON, K.S., COTTRELL, J.S., CORBETT, J.M., WHEELER, C.H., DUNN, M.J. AND PAPPIN, D.J. (1995). Identification of myocardial proteins from two-dimensional gels by peptide mass fingerprinting. *Electrophoresis*, 16, 308-316.
- TOL, G.J., FAL'SNAUGH, J.L., AKINYOSOYE, O., LACKLAND, H., WINTERCASH, P., VITORICA, F.J. AND STEIN, S. (1989). Amino acid analysis on polyvinylidene difluoride membranes. *Analytical Biochemistry*, 179, 50-55.
- TOVEY, E.R., FORD, S.A. AND BALDO, B.A. (1987). Protein blotting on nitrocellulose: some important aspects of the resolution and detection of antigens in complex extracts. *Journal of Biochemical and Biophysical Methods*, 14, 1-17.
- URWIN, V.E. AND JACKSON, P. (1993). Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: imaging of the fluorescent protein spot patterns using a cooled charge-coupled device. *Analytical Biochemistry*, 209, 57-62.
- VAN BOGELEN, R.A., HUTTON, M.E. AND NEIDHARDT, F.C. (1990). Gene-protein database

- of *Escherichia coli*. N-12; edition 3. *Electrophoresis*, 11, 1131-1166.
- VAN BOGELEN, R.A. AND NEIDHARDT, F.C. (1991). The gene-protein database of *Escherichia coli*; edition 4. *Electrophoresis*, 12, 955-994.
- VAN BOGELEN, R.A., SANKER, F., CLARK, R.L., BOGAN, J.A. AND NEIDHARDT, F.C. (1992). The gene-protein database of *Escherichia coli*; edition 5. *Electrophoresis*, 13, 101-1054.
- VANDEKERKHOVE, J., BALW, G., VANCOUVERNOLLE, K., HOVORE, B. AND CELIS, J. (1991). Comparative two-dimensional gel analysis and microsequencing identifies gelolin as one of the most prominent downregulated markers of transformed human fibroblast and epithelial cells. *Journal of Cell Biology*, 111, 95-102.
- VANFLETEREN, J.R., RAYNACKERS, J.G., VAN BUN, S.M. AND MEHLIS, L.A. (1992). Peptide mapping and microsequencing of proteins separated by SDS-PAGE after limited *in situ* hydrolysis. *BioTechniques*, 12, 550-557.
- VORM, O. AND MANN, M. (1994). Improved mass accuracy in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of peptides. *Journal of the American Society for Mass Spectrometry*, 5, 955-958.
- VORM, O., ROEPSTORFF, P. AND MANN, M. (1994). Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Analytical Chemistry*, 66, 3281-3287.
- WALLACE, A. AND SALLZ, H.P. (1992a). Ultramicrodetection of proteins in polyacrylamide gels. *Analytical Biochemistry*, 203, 27-34.
- WALLACE, A. AND SALLZ, H.P. (1992b). Beyond silver staining. *Nature*, 357, 608-609.
- WALSH, B.J., GOOLEY, A.A., WILLIAMS, K.L. AND BRETT, S.N. (1995). Identification of macrophage activation associated proteins by two-dimensional electrophoresis and microsequencing. *Journal of Leukocyte Biology*, 57, 507-512.
- WASINGER, V.C., CORDWELL, S.J., POLJAK, A., YAN, J.X., GOOLEY, A.A., WILKINS, M.R., DUNCAN, M., HARRIS, R., WILLIAMS, K.L. AND HUMPHREY-SMITH, I. (1995). Progress with Gene-Product Mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*, 16, In Press.
- WEITZHANDLER, M., KADLECEK, D., AVDALOVIC, N., FORTE, J.G., CHOW, D. AND TOWNSEND, R.R. (1993). Monosaccharide and oligosaccharide analysis of proteins transferred to polyvinylidene fluoride membranes after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Journal of Biological Chemistry*, 268, 5121-5130.
- WILKINS, M.R., PASQUALI, C., APPEL, R.D., LUO, K., GOLAZ, O., SANCHEZ, J.C., YAN, J.X., GOOLEY, A.A., HUGHES, G., HUMPHREY-SMITH, I., WILLIAMS, K.L. AND HOCHSTRASSER, D.F. (1995). From Proteins to Proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Submitted.
- WILKINS, M.R., LUO, K., APPEL, R.D., GOLAZ, O., PASQUALI, C., YAN, J.X., FARNSWORTH, V., CARTIER, P., HOCHSTRASSER, D.F., WILLIAMS, K.L. AND GOOLEY, A.A. (1996). Rapid protein identification using N-terminal sequence tagging and amino acid analysis (submitted).
- WIRTH, P.J., LUO, L.D., FUJIMOTO, Y., BISGAARD, H.C. AND OLSEN, A.D. (1991). The rat liver epithelial (RLE) cell protein database. *Electrophoresis*, 12, 921-934.
- WIRTH, P.J., LUO, L.D., BENJAMIN, T., HOANG, T.N., OLSEN, A.D. AND PARMALEE, D.C. (1993). The rat liver epithelial (RLE) cell nuclear protein database. *Electrophoresis*, 14, 1199-1215.
- WI, Y., LEMIKIN, P.F. AND UPTON, K. (1993). A fast spot segmentation algorithm for two-dimensional gel electrophoresis analysis. *Electrophoresis*, 14, 1351-1356.
- YAMAGUCHI, K. AND ASAKAWA, H. (1988). Preparation of colloidal gold for staining proteins electrotransferred onto nitrocellulose membranes. *Analytical Biochemistry*, 172, 104-107.
- YAMASHITA, K., IDEO, H., OHKURA, T., FUKUSHIMA, K., YIASA, I., OINO, K. AND TAKESHITA, K. (1993). Sugar chains of serum transferrin from patients with carbohydrate deficient glycoprotein syndrome. Evidence of asparagine-N-linked oligosaccharide transfer deficiency. *Journal of Biological Chemistry*, 268, 5783-5789.
- YATES, J.R. III, SPEICHER, S., GRIFFIN, P.R. AND HUNKAPILLER, T. (1993). Peptide mass maps: a highly informative approach to protein identification. *Analytical Biochemistry*, 214, 397-408.

The

RM D

Thorn
Hamilton
Wales
Private

Introduction

Proteins dealing type. The structure since the by Watson (1930) a possibly mechanism subtilis 1986; I commented protein the chain been in regulation apparel Kalisz, Barrell Give surprise attract subject stability enzymic Sherod

* Content:
Research 1264-872

Human cellular protein patterns and their link to genome DNA sequence data: usefulness of two-dimensional gel electrophoresis and microsequencing

JULIO E. CELIS,^{1,2} HANNE H. RASMUSSEN,¹ HENRIK LEFFERS,¹ PEDER MADSEN,¹ BENT HONORE,¹ BORBALA GESSER,¹ KURT DEJGAARD,¹ JOËL VANDEKERCKHOVE²

¹Institute of Medical Biochemistry and Human Genome Research Centre, Aarhus University, DK-8000 Aarhus, Denmark, and ²Laboratorium voor Fysiologische Chemie, Rijksuniversiteit Gent, Belgium

ABSTRACT Analysis of cellular protein patterns by computer-aided 2-dimensional gel electrophoresis together with recent advances in protein sequence analysis have made possible the establishment of comprehensive 2-dimensional gel protein databases that may link protein and DNA information and that offer a global approach to the study of the cell. Using the integrated approach offered by 2-dimensional gel protein databases it is now possible to reveal phenotype specific protein (or proteins), to microsequence them, to search for homology with previously identified proteins, to clone the cDNAs, to assign partial protein sequence to genes for which the full DNA sequence and the chromosome location is known, and to study the regulatory properties and function of groups of proteins that are coordinately expressed in a given biological process. Human 2-dimensional gel protein databases are becoming increasingly important in view of the concerted effort to map and sequence the entire genome.—Celis, J. E.; Rasmussen, H. H.; Leffers, H.; Madsen, P.; Honoré, B.; Gesser, B.; Dejgaard, K.; Vandekerckhove, J. Human cellular protein patterns and their link to genome DNA sequence data: usefulness of two-dimensional gel electrophoresis and microsequencing. *FASEB J.* 5: 2200-2208; 1991.

Key Words: human protein patterns • 2-dimensional gel protein databases • gene expression • microsequencing • cDNA cloning • linking protein and DNA information • genome mapping and sequencing

PROTEINS SYNTHESIZED FROM information contained in the DNA orchestrate most cellular functions. The total number of proteins synthesized by a typical human cell is unknown although current estimates range from 3000 to 6000. Of these, as many as 70% may perform household functions and are expected to be shared by all cell types irrespective of their origin. There are many different cell types in the human body with perhaps 30,000 to 50,000 proteins expressed in the organism as a whole judged from the fact that about 3% of the haploid genome correspond to genes. Today only a small fraction of the total set of proteins has been identified, and little is known about the protein patterns of individual cell types or their variation under physiological and abnormal conditions.

For the past 15 years, high resolution 2-dimensional gel electrophoresis has been the technique of choice to determine the protein composition of a given cell type and for monitoring changes in gene activity through quantitative and qualitative analysis of the thousands of proteins that orchestrate various cellular functions (refs 1-6 and references

therein). The technique originally described by O'Farrell separates proteins in terms of their isoelectric point (pI) and molecular weight. Usually one chooses a condition of interest and the cell reveals the global protein behavioral response as all detected proteins can be analyzed both qualitatively and quantitatively in relation to each other. At present, most available 2-dimensional gel techniques (regular gel format) can resolve between 1000 and 2000 proteins from a given mammalian cell type, a number that corresponds to about 2 million base pairs of coded DNA. Less abundant proteins can be detected by analyzing partially purified cellular fractions.

Two-dimensional gel electrophoresis has been widely applied to analysis of cellular protein patterns from bacteria to mammalian cells (refs 1-6, and references therein). In spite of much work, however, information gathered from these studies has not reached the scientific community in its fullness because of lack of standardized gel systems and the lack of means for storing and communicating protein information. Only recently, because of the development of appropriate computer software (7-13), has it been possible to scan gels, assign numbers to individual proteins, and store the wealth of information in quantitative and qualitative comprehensive 2-dimensional gel protein databases (4, 14-23), i.e., those containing information about the various properties (physical, chemical, biological, biochemical, physiological, genetic, immunological, architectural, etc.) of all the proteins that can be detected in a given cell type. Such integrated 2-dimensional gel protein databases offer an easy and standardized medium in which to store and communicate protein information and provide a unique framework in which to focus a multidisciplinary approach to study the cell. Once a protein is identified in the database, all of the information accumulated can be easily retrieved and made available to the researcher. In the long run, protein databases are expected to foster a wide variety of biological information that may be instrumental to researchers working in many areas of biology—among others, cancer and oncogene studies, differentiation, development, drug development and testing, genetic variation, and diagnosis of genetic and clinical diseases (Fig. 1).

The approach using systematic 2-dimensional gel protein analysis has recently gained a new dimension with the advent of techniques to microsequence major proteins recorded

¹To whom correspondence should be addressed, at: Institute of Medical Biochemistry and Human Genome Research Centre, Ole Worms Alle, Bldg. 170, University Park, DK-8000 Aarhus C, Denmark.

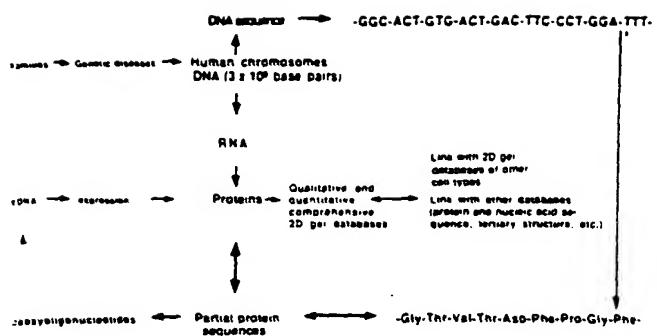


Figure 1. Interface between partial protein sequence databases, comprehensive 2-dimensional gel databases, and the human genome sequencing project. Appropriate software is required to compare protein and DNA sequences. In general, although the inference of a protein's sequence from the DNA sequence (thick arrow) is direct and unambiguous, the DNA sequence can only be inferred approximately from the protein sequence (thin arrow) and cloning of the gene requires either a cDNA or the requisite group of oligonucleotide probes deduced from the partial amino acid sequence. Modified from ref. 6.

in the databases (refs 24-42 and references therein). Partial protein sequences can be used to search for protein identity as well as to prepare specific DNA probes for cloning as-yet uncharacterized proteins (Fig. 1). As these sequences can be stored in the database (see for example Fig. 2H), they offer a unique opportunity to link information on proteins with the existing or forthcoming DNA sequence data on the human genome (Fig. 1) (20, 36, 39).

Using the integrated approach offered by comprehensive 2-dimensional gel databases (Fig. 1), it will be possible to identify phenotype-specific proteins; microsequence them and store the information in the database; search for homology with previously characterized proteins; clone the cDNAs; assign partial protein sequences to genes for which the full DNA sequence and the chromosome location are known, and study the regulatory properties and function of groups of proteins (pathways, organelles, etc.) that are coordinately expressed in a given biological process. Comprehensive 2-dimensional gel protein databases will depict an integrated picture of the expression levels and properties of the thousands of protein components of organelles, pathways, and cytoskeletal systems in both physiological and abnormal conditions and are expected to lead to identification of new regulatory networks in different cell types and organisms. In the future, 2-dimensional gel protein databases may be linked to each other as well as to national and international specialized databanks on nucleic acid and protein sequences, protein structures, NMR experimental data, complex carbohydrates, etc.

A few 2-dimensional gel protein databases that are accessible in a computer form have been published in extenso; these correspond to the protein-gene database of *Escherichia coli* K-12 developed by Neidhardt and colleagues (14, 23), the rat REF 52 database established by Garrels and co-workers at Cold Spring Harbor (18, 22), and a few human databases (transformed amnion cells [15, 20], normal embryonal lung MRC-5 fibroblasts [17, 21], keratinocytes [19] and peripheral blood mononuclear cells [15]) developed in Aarhus. Given space limitations and to keep this review in focus, we will concentrate on the computerized analysis of human cellular 2-dimensional gel patterns, and in particular on the steps involved in establishing comprehensive 2-dimensional gel databases that can link protein and DNA information.

MAKING AND MANAGING A COMPREHENSIVE 2-DIMENSIONAL GEL DATABASE OF HUMAN CELLULAR PROTEINS

The first step in making a comprehensive 2-dimensional gel protein database is to prepare a synthetic image (digital form of the gel image) of the gel (fluorogram, Coomassie blue or silver stained gel) to be used as a standard or master reference. This can be done with laser scanners, charge couple device (CCD)² array scanners, television cameras, rotating drum scanners, and multiwire chambers (13). Computerized analysis systems for spot detection, quantitation, pattern matching, and data handling (access and retrieval of information, database making) have been described in the literature (ELSIE [43], GELLAB [11], HERMeS [44], MELANIE [10], QUEST [9], and TYCHO [8]) and some are available commercially (PDQUEST, Protein Database Inc., Huntington, N.Y.; KEPLER, Large Scale Biology, Rockville, Md.; Visage, BioImage Corporation, Ann Arbor, Mich.; Gemini, Joyce Loeb, Gateshead; Microscan 1000, Technology Resources Inc., Nashville, Tenn. and MasterScan, Billerica, Mass.). Unfortunately, most of these systems are incompatible with one another and their advantages and disadvantages have been discussed by Miller (13).

In our work station in Aarhus, fluorograms are scanned with a Molecular Dynamics laser scanner and the data are analyzed using the PDQUEST II software (Protein Databases Inc.) (12) running on a spark station computer 4100 FC-8-P3 from SUN Microsystems, Inc. The scanner measures intensity in the range of 0-2.0 absorbance. A typical scan of a 17 x 17 cm fluorogram takes about 2 min. Steps in image analysis include: initial smoothing, background subtraction, final smoothing, spot detection, and fitting of ideal Gaussian distribution to spot centers. Spot intensity is calculated as the integration of a fitted Gaussian. If calibration strips containing individual segments of a known amount of radioactivity are used, it is possible to merge multiple exposures of the sample image into a single data image of greater dynamic range. Once the synthetic image is created it can be stored on disk and displayed directly on the monitor. Functions that can be used to edit the images include: cancel (for example, to erase scratches that may have been interpreted as spots by the computer; cancel streaks or low dpm spots), combine (sometimes a spot may be resolved into several closely packed spots), restore, uncombine, and add spot to the gel. The process is time consuming—about 1-1/2 day per image. Edited standard images can be matched to other synthetic images. Figure 2A shows a portion of a standard synthetic image (IEF) of a fluorogram of [³⁵S]methionine labeled cellular proteins from human AMA cells (master database) (20). Images can be displayed either in black and white (resembling the original fluorograms) or in color (other images in Fig. 2), depending on the need. As shown in Fig. 2B, each polypeptide is assigned a number by the computer, which facilitates the entry and retrieval of qualitative and quantitative information for any given spot in the gel (20). The standard image can be matched automatically by the computer to other standard or reference gels (Fig. 2C, matching of AMA cellular proteins [left] to MRC-5 proteins [right]) provided a few landmark spots are given manually as reference (indicated with a + in Fig. 2C) to initiate the process.

²Abbreviations: CCD, charge couple device; PCNA, proliferating cell nuclear antigen; HPLC, high performance liquid chromatography.

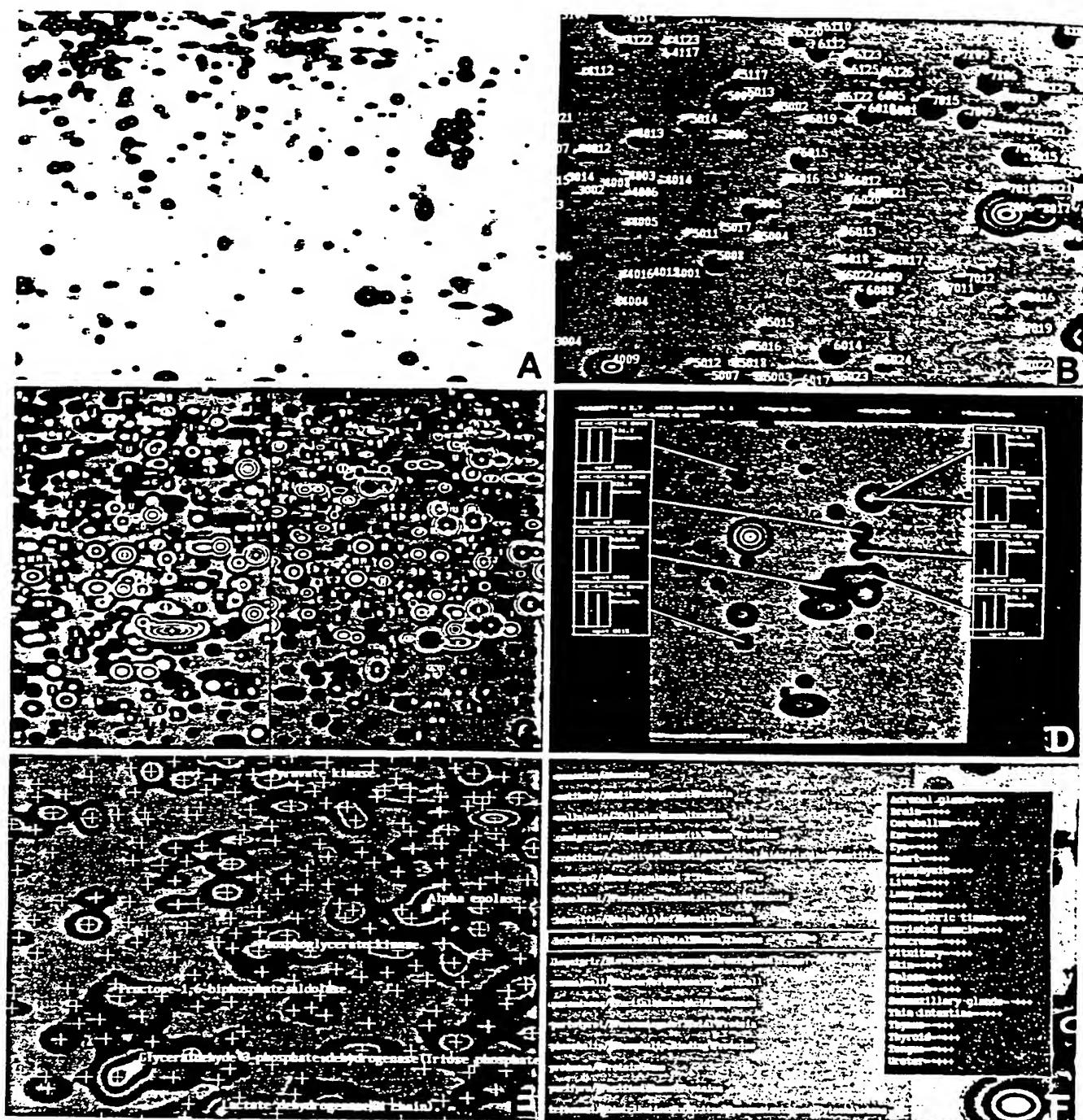
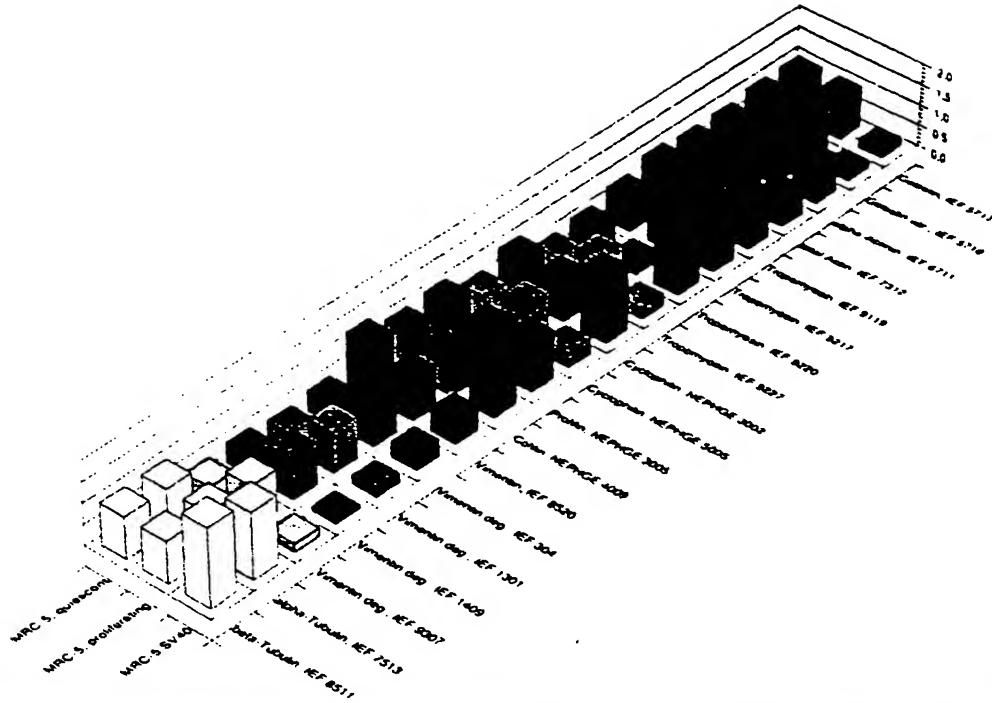
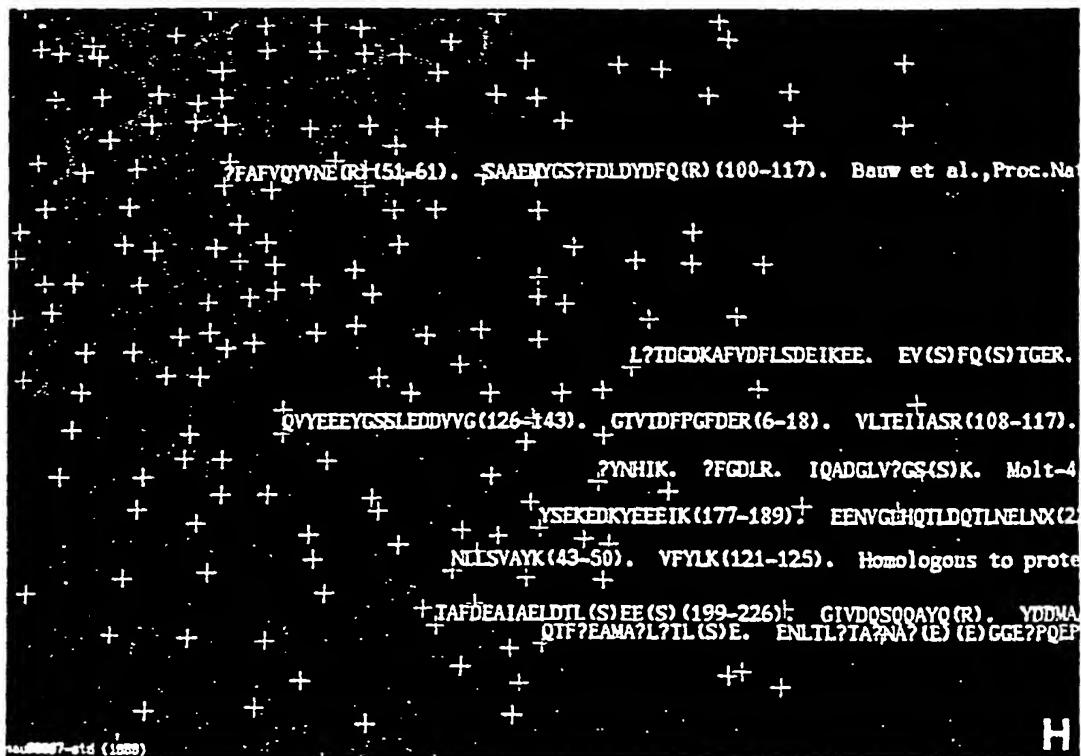


Figure 2. *A)* Synthetic image of a fraction of an IEF gel of the master image of AMA cellular proteins. *B)* As in *A* but showing numbers assigned to each spot. *C)* Comparison of AMA (left) and normal human embryonal lung MRC-5 fibroblasts (right) IEF protein patterns. Matched proteins are indicated by a + or by the same letters in both gels. Once a protein is matched, information contained in the various categories available in the master AMA database can be transferred. *D)* Synthetic image of a fraction of an IEF fluorogram of [³⁵S]methionine labeled proteins from normal human MRC-5 fibroblasts. The histograms show levels of synthesis of a few proteins in MRC-5 (left bar) and SV40 transformed MRC-5 (right bar) fibroblasts. *E)* Polypeptides that contain information under the category glycolytic pathway. *F)* The function peruse annotation for spot allows the operator to inquire about categories and information available for a given protein. *G)* Relative abundance of cytoskeletal and cytoskeletal-related proteins in quiescent, proliferating, and SV40-transformed MRC-5 fibroblasts. *H)* Polypeptides that contain information under the category partial amino acid sequences.



G



H

The automatic matching process that has been described in detail by Garrels et al. (12) takes about 5 min. Matched proteins are indicated with the same letters in both gels (Fig. 2C). The usefulness of this function is emphasized by the fact that data accumulated on common household proteins can be easily transferred to any other human cellular cell type whose 2-dimensional gel cellular protein pattern is matched

to our standard AMA 2-dimensional gel protein image. Alternatively, if the standard gel is part of a matchset (set of gels in a given experiment) it can be used as a linker gel to compare, for example, the quantitative values of a given protein throughout the experiment (see Fig. 2D; levels of some proteins in normal and SV40 transformed human MRC-5 fibroblasts) or with other standard images in different sets of

cross-matched experiments (18, 22).

Once a standard map of a given protein sample is made, one can enter qualitative annotations to make a reference database. Our master 2-dimensional gel database of transformed human amnion cell (AMA) proteins (20) lists 3430 polypeptides of which 2592 correspond to cellular components, having pI's ranging from 4 to 13 and molecular weights between 8.5 and 230 kDa. The most abundant proteins in the database correspond to total actin (3.87% of total protein; about 90 million molecules per cell) while the lesser abundant of the recorded polypeptides are present in the vicinity of 5000 molecules per cell. Some annotation categories we are using to establish the master AMA database include: 1) protein identification (comigration with purified proteins, 2-dimensional immunoblotting, microsequencing); 2) amounts (total amounts and levels of synthesis); 3) subcellular localization (nuclear, cytoskeletal, membrane, membrane receptors, specific organelles, etc.); 4) antibodies; 5) posttranslational modifications (phosphorylation, glycosylation, methylation etc.); 6) microsequencing; 7) cell cycle specificity (specific variations in levels of synthesis and amount); 8) regulatory behavior (effect of hormones, growth factors, heat shock, etc.); 9) rate of synthesis in normal and transformed cells (proliferation sensitive proteins, cell cycle specific proteins, oncogenes, components of the pathway (or pathways) that control cell proliferation); 10) function (mainly from comigration with proteins of known function); 11) sets of proteins that are coordinately regulated (hierarchy of controls, differential gene expression in various cells, etc.); 12) cDNAs (cloned cDNAs); 13) proteins that are specific to a given disease (systematic comparison of protein patterns of fibroblast proteins from healthy and diseased individuals); 14) expression and exploitation of transfected cDNAs; 15) pathways (metabolic, others); 16) gene localization (genetic and physical); 17) effect of microinjected antibody on patterns of protein synthesis; and 18) secreted proteins.

Information entered for any spot in a given annotation category can be easily retrieved by asking the computer to display the information on the color screen. For example, Fig. 2E shows a synthetic image of a NEPHGE gel (master AMA database) displaying the information contained under the entry glycolytic pathway. Alternatively, one can use the function peruse annotations for spot to directly ask the computer to list all the entries available for a particular protein. By clicking the mouse in a given entry (in this case, presence in fetal human tissues) it is possible to take a quick look at the information in that particular entry (Fig. 2F).

A major obstacle encountered in building comprehensive 2-dimensional gel protein databases is identifying the large number of proteins separated by this technology. In our databases (20, 21), known proteins are identified by one or a combination of the following procedures: 1) comigration with known proteins, 2) 2-dimensional gel immunoblotting using specific antibodies, and 3) microsequencing of Coomassie Brilliant Blue stained human proteins recovered from dried 2-dimensional gels (see next section). Protein identification by means of microsequencing may be difficult, as individual protein members of families with short peptide differences may escape detection. In the gene-protein database of *E. coli* K-12 (14, 23), another major 2-dimensional gel database available at present, proteins are being identified by a wider range of tests that include comigration with purified proteins; genetic criterion (deletion, insertion, frameshift, nonsense, missense, regulatory), plasmid-bearing strains and in vitro synthesis of protein; selective labeling (methylation, phosphorylation); peptide map similarity; and physiological criterion and selective derivatization.

So far we have received nearly 550 antibodies from laboratories all over the world and these are being systematically tested by 2-dimensional gel immunoblotting for antigen determination. Similarly, purified proteins and organelles provided by several laboratories have greatly aided identification of unknown proteins (20, 21). We routinely request antibodies and protein samples and promise the donors to make available all the information we may have accumulated on that particular protein. For example, Table 1 lists entries available for Lipocortin V (IEF SSP 8216), also known as annexin V, VAC- α , endonexin II, renocortin, chromobindin-5', anticoagulant protein, PAP-I, γ -calcimedin, IBC, calphobindin, and anchorin CII.

As mentioned previously, one distinct advantage of 2-dimensional gel electrophoresis is the possibility of studying quantitative variations in cellular protein patterns that may lead to identification of groups of proteins that are expressed coordinately during a given biological process. Quantitation, however, is not an easy task as reflected by the lack of published data on global cellular protein patterns. We believe this is partly due to difficulties in obtaining sets of gels that are suitable for computer analysis (streaking, material remaining at the origin, etc.) as well as to limitations (laborious editing time, need of calibration strips to merge images, limited dynamic range, etc.) in the computer analysis systems available at the moment. Perhaps the most advanced quantitative studies published so far using computer analysis have been carried out by Garrels and co-workers (18, 22). In particular, these investigators have established a quantitative rat protein database (18, 22) designed to study growth control (proliferation, growth inhibitors, and stimulation) and transformation in well-defined groups of cell lines obtained by transformation of rat REF52 cells with SV40, adenovirus, and the Kirsten murine sarcoma virus. These studies have revealed clusters of proteins induced or repressed during growth to confluence as well as groups of transformation-sensitive proteins that respond in a differential fashion to transformation by DNA and RNA viruses. A most interesting feature of this quantitative database is the discovery of a group of coregulated proteins that show similar expression patterns as the cell cycle-regulated DNA replication protein known as proliferating cell nuclear antigen (PCNA)/cyclin (45).

In our human databases, most quantitations have been carried out by estimating the radioactivity contained in the polypeptides by direct counting of the gel pieces in a scintillation counter (20, 21). Up to 700 proteins can be cut out through appropriate exposed films in a period of time comparable to that required for editing a synthetic image. Manual quantitation of this large number of spots is difficult without the assistance of a master reference image and a numbering system that can be used to identify the spots. Using this approach, we have recorded quantitative changes in the relative abundance of 592 [35 S]methionine-labeled proteins synthesized by quiescent, proliferating, and SV40 transformed human embryonic lung MRC-5 fibroblasts (21). Some data concerning cytoskeletal and cytoskeletal-related proteins are presented in Fig. 2G. Our studies as well as those of Garrels and co-workers (18, 22) may in the long run help define patterns of gene expression that are characteristic of the transformed state.

OTHER 2-DIMENSIONAL GEL PROTEIN DATABASES

As mentioned previously there are other 2-dimensional gel databases available in computer form that have been pub-

TABLE 1. Some entries for lipocortin V in the human A.M.A. 2-dimensional gel protein database

Entries for lipocortin V (IEF SSP 8216)	Information entered
1. Protein name	Lipocortin V, renocortin, chromobindin-5, endonexin I, anticoagulant protein PAP-I, VAC- α , 35- γ -calcimedin, IBC, calphobindin I, anchorin CII, annexin V
2. Percentage of total protein	0.110% (about 2,800,000 molecules per cell)
3. Apparent molecular weight (mr)	33.3 kDa
4. Isoelectric point (pl)	4.76
5. Method (or methods) of identification	Microsequencing, 2-dimensional immunoblotting, Comigration
6. Credit to investigators that aided in identification	G. Bauw, J. Vandekerckhove, and colleagues, Rijksuniversiteit Gent; B. Pepinsky, BIOGEN, Cambridge; N.G. Ahn, University of Washington
7. Antibody against protein	Polyclonal (rabbit, antibody no. 20), B. Pepinsky, BIOGEN, Cambridge
8. Comigration with human proteins	Lipocortin V, N.G. Ahn, Howard Hughes Medical Institute, Washington University
9. Cellular localization	Subcortical membrane
10. Calcium/phospholipid-dependent membrane proteins	Lipocortin V
11. Function	Regulation of various aspects of inflammation, immune response, blood coagulation and differentiation
12. Partial amino acid sequence	GTVTDFPGFDER (7-18), VLTEIIASR (109-117), QVYEEEYGSSLEDDVVVG (127-143), ?GTDEEKFITIFGT(R) (187-201)
13. cDNA sequence	Known, R. Blake et al., <i>J. Biol. Chem.</i> 263, 10799-10811, 1988 (pl = 4.76 from translated sequence)
14. Levels in fetal human tissues	Adrenal glands = + + + ; brain = - - - ; cerebellum = + + - ; ear = + + + ; eye = - - - ; heart = - - - ; hypophysis = + + + ; liver = + + + ; lung = + + + ; meninges = + + + ; mesonephric tissue = + + + ; striated muscle = + + + ; pancreas = + + + ; skin = + + + ; spleen = + + + ; stomach = + + + ; submandibular gland = + + + ; small intestine = - - - ; thymus = - - - ; thyroid gland = + + + ; tongue = + + + ; ureter = + + +
15. Levels in quiescent, proliferating, and transformed MRC-5 fibroblasts	Q (quiescent) = 1.1; P (proliferating) = 1.0; T (SV+0 transformed) = 0.3
16. Distribution in Triton supernatant and cytoskeletons	Mainly supernatant

lished in extenso: these correspond to the *E. coli* K-12 protein-gene database (14, 23) and to the rat REF52 database (18, 22).

The *E. coli* K-12 cellular protein-gene database is perhaps the most complete of all databases reported so far and eventually it should trace each protein back to its structural gene. Information contained in this database includes: gene/protein name (protein name, EC number, gene name); 2-dimensional gel spot designations (x-y coordinates from reference gels, alphanumeric designation); genetic information (linkage map location, physical map location, Genebank code, sequence reference, location on Kohara clones); biochemical information (molecular weight, pl, number of residues of each amino acid, mole percent of each amino acid, total number of amino acids in a polypeptide), and regulatory information (cellular level of protein in different media and different temperature, member of regulon, member of stimulon). Major advances of this database are envisaged in the future in view of the eminent sequencing of

the whole *E. coli* genome as well as the development of improved methods to express cloned genes.

The rat REF52 2-dimensional gel protein database lists about 1600 proteins that have been recorded using the QUEST analysis system (18, 22). Included in this quantitative database are 1) protein names (cytoskeletal and heat shock proteins as well as various nuclear, mitochondrial, and cytoplasmic proteins), 2) annotations (subcellular localization, modification, recognition by specific antibodies, coprecipitation, NH₂-terminal sequence, cross-reference to protein sequence information and references to the literature), 3) protein sets (cytoskeletal proteins, phosphoproteins, sets of proteins with PCNA/cyclin-like properties, etc.) and 4) general quantitative data (protein synthesis during growth of normal REF52 cells to confluence and quiescence, and after restimulation of growth-inhibited cells).

In addition to the 2-dimensional gel databases mentioned so far there are several smaller cellular databases being established in human (normal human diploid fibroblasts, lym-

phocytes, leukocytes, leukemic cells) mouse (NIH/3T3 cells, T lymphocytes), *Aplysia*, yeast (*Saccharomyces cerevisiae*), plants (wheat, barley, sorghum), and *Euglena*. Databases of tissue protein (brain, whole mouse, liver) and body fluid proteins (plasma proteins, cerebrospinal fluid, urine, and milk) are being established in several laboratories. The reader is directed to the review by Celis et al. (4) for details and references concerning these databases.

MICROSEQUENCING HAS ADDED A NEW DIMENSION TO COMPREHENSIVE 2-DIMENSIONAL GEL DATABASES: A DIRECT LINK BETWEEN PROTEINS AND GENES

The development of highly sensitive amino acid gas-phase or liquid-phase sequenators (24), together with the establishment of efficient protein and peptide sample preparation methods, has opened the possibility to perform a systematic sequence analysis of proteins resolved by 2-dimensional gel electrophoresis. Indeed, generated pieces of protein sequences can be used to search for protein identity (comparison with available sequences stored in databanks) as well as for preparing specific DNA probes for cloning of as yet uncharacterized proteins (Fig. 1). In addition, partial protein sequences can be stored in 2-dimensional gel databases (for example, see Fig. 2H) and offer a unique link between proteins and genes (Fig. 1).

In the early 1970s gel electrophoresis was used to purify proteins for sequencing purposes (reviewed by Weber and Osborn in ref 25). Proteins were recovered by diffusion and sequenced by the manual dansyl-Edman degradation at the nanomole level. This technique was further refined by using electro-elution to recover proteins and by miniaturizing the system (26). This method has been used extensively, but showed increasing drawbacks (low yields, protein samples contaminated by free amino acids, and NH₂-terminal blocking) as the amounts of handled protein gradually became smaller (e.g., at the 10 picomol level).

Most of the problems referred to above have been minimized with the introduction of protein-electroblotting procedures (27-32). When proteins are blotted on chemically inert membranes, it is possible to sequence the immobilized proteins directly without additional manipulations. Thus, depending on the amount of bound protein and its nature, this direct sequencing procedure generally yields NH₂-terminal sequences containing 10-40 residues. As such, this technique was used to identify, by their NH₂-terminal sequences, differentially expressed major proteins from total cellular extracts separated on 2-dimensional gels. A major difficulty encountered in this procedure is the occurrence of frequent artefactual blockage of the proteins. Several studies suggest that this phenomenon is mainly due to reaction with contaminants (particularly unpolymerized acrylamide present in the gel) and to a high dilution of the protein (low concentration of the protein per unit membrane surface). In addition to this primarily technical problem, many proteins are blocked in vivo by acylation or by a pyrrolidone carboxylic acid cap.

The problem of partial or complete NH₂-terminal blockage can be circumvented by generating internal amino acid sequences. This is achieved by fragmenting the protein present in the gel (gel *in situ* cleavage) or by cleaving it while bound to the membrane (membrane *in situ* cleavage) (33-35). In both cases, proteins are either cleaved in a restricted way (e.g., by limited enzymatic digestion or by using restriction chemical cleavage conditions) or fragmented into smaller peptides.

Of the different combinations examined, we had good results by using exhaustive proteolytic digestion on membrane-immobilized proteins. This method has been described for Ponceau red-stained proteins on nitrocellulose-blots (34), for Amido-black-stained Immobilon-bound proteins, and for fluorescamine-detected proteins on glass fiber membranes (35). The proteases used (trypsin, chymotrypsin, or pepsin) cleave at multiple sites, generating small peptides that elute from the blot into the digestion buffer from which they are purified by reversed-phase high performance liquid chromatography (HPLC) before being sequenced individually. Although each of these manipulations could be expected to result in a reduced yield of final sequence information, we were surprised that the peptides could be sequenced with high efficiency. In our hands, this approach could be routinely applied to gel-purified proteins available in amounts ranging from 5 to 10 µg, and often yielded sequence information covering more than 30% of the total protein. As membrane-immobilized proteins are not homogeneously digested, but rather show protease sensitivity next to resistant regions, the number of peptides generated is much lower than expected from the number of potential cleavage sites. Consequently, HPLC peptide chromatograms are less complex and most peptides can be recovered in pure form.

As only limited amounts of a protein mixture can be loaded on a 2-dimensional gel, proteins of interest are often obtained in yields insufficient for the currently available sequencing technology. More material can be obtained by enriching for a certain subcellular fraction (purified cell organelles) or by exploiting affinity (dyes, metals, drugs, etc) or hydrophobic properties of proteins before gel analysis. All of the sequencing results accumulated so far in the human protein database (20) (a few are shown in Fig. 2H) have been obtained from analysis of protein spots collected from 2-dimensional gels that had been stained with Coomassie blue according to standard procedures and dried for storage. Proteins are recovered from the collected gel pieces by a protein-elution-concentration device, combined with gel electrophoresis and electroblotting. Details of this technique have been reported in a previous communication (42) and a brief outline is given below.

Combined gel pieces are allowed to swell in gel sample buffer (a total volume of 1.5 ml). The gel pieces combined with the supernatant are then collected into a large slot made in a new gel. The slot is further filled with Sephadex G-10 equilibrated in gel sample buffer. During consecutive gel electrophoresis, most of the electrical current passes on the side of the slot instead of passing through the slot. This results in both a vertical stacking and horizontal contraction of the protein band. With this device the protein is efficiently eluted from the gel pieces and concentrated from a large volume into a narrow spot. The highly concentrated (about 5 mm²) protein spot is then electroblotted on PVDF-membranes, stained with Amido black, and *in situ* digested with trypsin. The peptides generated during digestion elute from the membrane into the supernatant, and can be separated by narrow bore reversed-phase HPLC and collected individually for sequence analysis.

Using this and previous procedures (37, 39, 42), we have so far analyzed 70 protein spots collected from 2-dimensional gels (20, and unpublished observations) (see for example Fig. 2H). The sequence information amounts to 2100 allocated residues corresponding to an average of 30 residues per protein spot. So far we have made cDNAs of many of the unknown proteins that have been microsequenced, and a substantial number has been cloned and sequenced. All available information indicates that it may be possible to obtain partial sequence information from most of

the proteins that can be visualized by Coomassie Brilliant Blue staining.

Partial protein sequences are stored in the database as displayed in Fig. 2H, and it should be possible in the near future to interface this information with forthcoming DNA sequence data from the human genome project. In the long run, as the human genome sequences become available it will be possible to assign partial protein sequences to genes for which the full DNA sequence and chromosomal location are known (Fig. 1).

SUMMARY

The studies presented in this brief review are intended to demonstrate the usefulness of computer-aided 2-dimensional gel electrophoresis and microsequencing to analyze cellular protein patterns, and to link protein and DNA information. As more information is gathered worldwide, comprehensive databases will depict an integrated picture of the expression levels and properties of the thousands of proteins that orchestrate most cellular functions.

Clearly, databases allow easy access to a large body of data and provide an efficient medium to communicate standardized protein information. In the future, databases will foster a wide variety of biological information that can be used to support collaborative research projects in basic and applied biology as well as in clinical research (2, 5, 46). Once a protein is identified in a particular database all the information gathered on it can be made available to the scientist. However, many problems must be solved before protein databases become of general use to the scientific community. A most urgent one is to promote standardization of the gel running conditions so that data produced in a given laboratory may be used worldwide. Surprisingly, the gel running technology as it stands today is still a craftsmanship art.

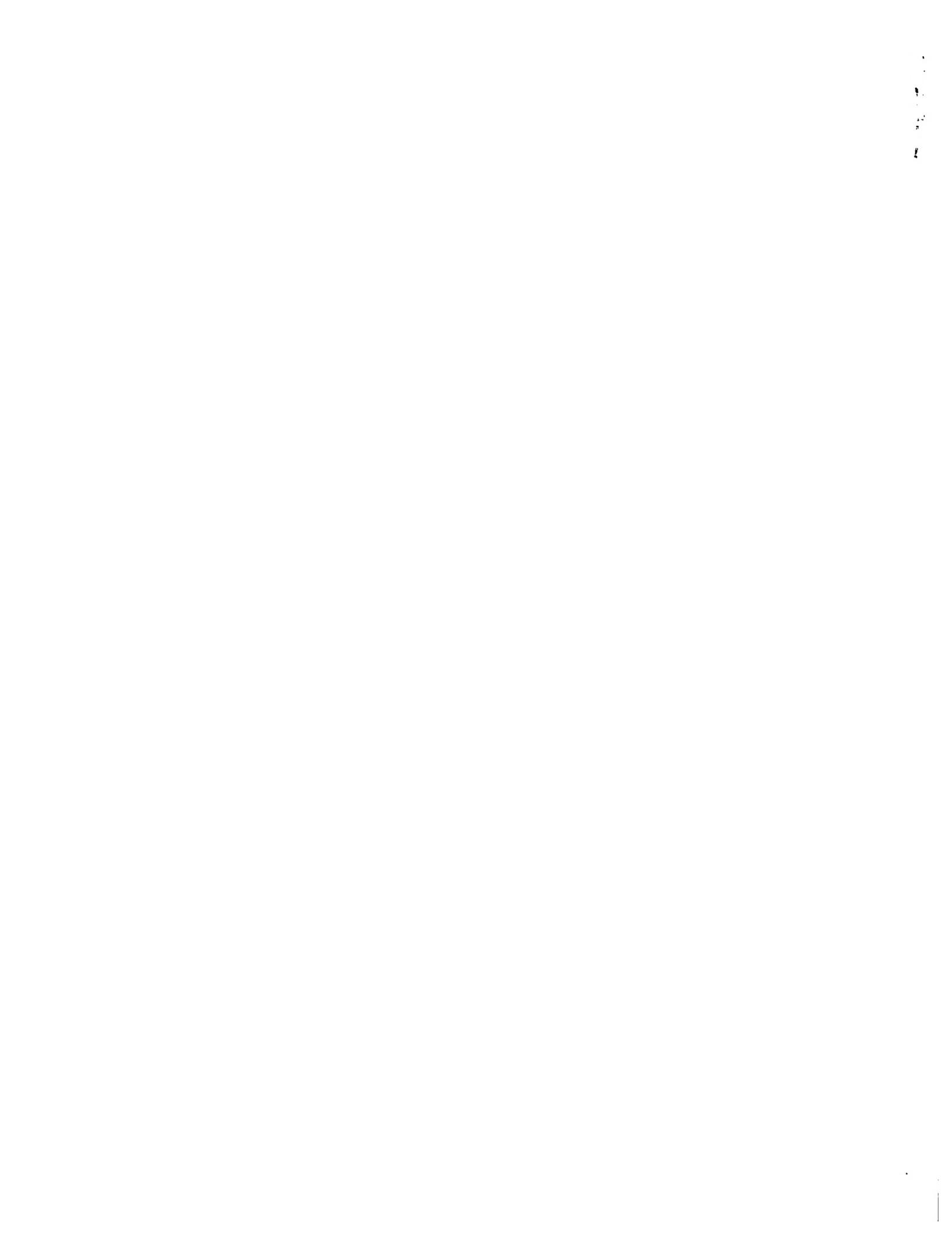
Finally, comprehensive, computerized databases of proteins, together with recently developed techniques to microsequence proteins, offer a new dimension to the study of genome organization and function (Fig. 1). In particular, human protein databases may become increasingly important in view of the concerted effort to map and sequence the entire human genome. This formidable task is expected to dominate biological research in the next decades. [FJ]

We would like to thank S. Himmelstrup Jørgensen for typing the manuscript and O. Sønderskov for photography. Work in the authors' laboratories was supported by grants from the Danish Biotechnology Programme, the Danish Cancer Foundation, and the Commission of the European Communities.

REFERENCES

1. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007-4021
2. Special Issue: Two-dimensional gel electrophoresis. *Clin. Chem.* **28**, 1982
3. Celis, J. E., and Bravo, R., eds. (1984) *Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications*. Academic, New York
4. Celis, J. E., Madsen, P., Gesser, B., Kwee, S., Nielsen, H. V., Rasmussen, H. H., Honoré, B., Leffers, H., Ratz, G. P., Basse, B., Lauridsen, J. B., and Celis, A. (1989) Protein databases derived from the analysis of two-dimensional gels. In *Advances in Electrophoresis* (Chrambach, C., ed) VCH, Weinheim, Germany
5. Special Issue: Two-dimensional gel electrophoresis in cell biology. (Celis, J. E., ed) *Electrophoresis* **11**, 1990
6. Celis, J. E., Honoré, B., Bauw, G., and Vandekerckhove, J. (1990) Comprehensive computerized 2D gel protein databases offer a global approach to the study of the mammalian cell. *BioEssays* **12**, 93-98
7. Garrels, J. I. (1983) Two-dimensional gel electrophoresis and computer analysis of proteins synthesized by cloned cell lines. *Methods Enzymol.* **100**, 411-423
8. Anderson, N. L., Hotimann, J. P., Gemmel, A., and Tavor, S. (1984) Global approaches to the quantitative analysis of gene-expression patterns observed by two-dimensional gel electrophoresis. *Clin. Chem.* **30**, 2031-2036
9. Garrels, J. I., Farrar, J. T., and Burwell, C. B. (1984) The Quest system for computer-analyzed two-dimensional electrophoresis of proteins in *Two-Dimensional Gel Electrophoresis of Proteins: Methods and Applications* (Celis, J. E., and Bravo, R., eds) pp. 37-91. Academic, New York
10. Vincens, P., and Tarroux, P. (1988) Two-dimensional electrophoresis computerized processing. *Int. J. Biochem.* **20**, 499-509
11. Appel, R., Hochstrasser, D., Roch, C., Funk, M., Müller, A. F., and Pellegrini, C. (1988) Automatic classification of two-dimensional gel electrophoresis pictures by heuristic clustering analysis: a step toward machine learning. *Electrophoresis* **9**, 136-142
12. Lemkin, P. F., and Lester, E. P. (1989) Database and search techniques for two-dimensional gel protein data: a comparison of paradigms for exploratory data analysis and prospects for biological modeling. *Electrophoresis* **10**, 122-140
13. Miller, M. J. (1989) Computer-assisted analysis of two-dimensional gel electrophoretograms. *Adv. Electrophoresis* **3**, 182-217
14. Phillips, T. D., Vaughn, V., Bloch, P. L., and Neidhardt, F. C. (1987) In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology; Gene-Protein Index of Escherichia coli K-12*, 2 ed. (Neidhardt, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaeffer, M., and Umbarger, H. E., ed) pp. 919-966. American Society for Microbiology, Washington, D.C.
15. Celis, J. E., Ratz, G. P., Celis, A., Madsen, P., Gesser, B., Kwee, S., Madsen, P. S., Nielsen, H. V., Yde, H., Lauridsen, J. B., and Basse, B. (1988) Towards establishing comprehensive databases of cellular proteins from transformed human epithelial amniotic cells (AMA) and normal peripheral blood mononuclear cells. *Leukemia* **9**, 561-601
16. Special Issue: Protein databases in two-dimensional electrophoresis. (Celis, J. E., ed) *Electrophoresis* **2**, 1989
17. Celis, J. E., Ratz, G. P., Madsen, P., Gesser, B., Lauridsen, J. B., Brogaard-Hansen, K. P., Kwee, S., Rasmussen, H. H., Nielsen, H. V., Crüger, D., Basse, B., Leffers, H., Honoré, B., Möller, O., and Celis, A. (1989) Computerized, comprehensive databases of cellular and secreted proteins from normal human embryonic lung MRC-5 fibroblasts: identification of transformation and/or proliferation sensitive proteins. *Electrophoresis* **10**, 76-115
18. Garrels, J. I., and Franz, B. R. (1989) The REF52 protein database. Methods of database construction and analysis using the Quest system and characterizations of protein patterns from proliferating and quiescent REF52 cells. *J. Biol. Chem.* **264**, 5283-5298
19. Celis, J. E., Crüger, D., Kiil, J., Dejgaard, K., Lauridsen, J. B., Ratz, G. P., Basse, B., Celis, A., Rasmussen, H. H., Bauw, G., and Vandekerckhove, J. (1990) A two-dimensional gel protein database of noncultured total normal human epidermal keratinocytes: identification of proteins strongly up-regulated in psoriatic epidermis. *Electrophoresis* **11**, 242-254
20. Celis, J. E., Gesser, B., Rasmussen, H. H., Madsen, P., Leffers, H., Dejgaard, K., Honoré, B., Olsen, E., Ratz, G., Lauridsen, J. B., Basse, B., Mouritzen, S., Hellerup, M., Andersen, A., Walbum, E., Celis, A., Bauw, G., Puype, M., Van Damme, J., and Vandekerckhove, J. (1990) Comprehensive two-dimensional gel protein databases offer a global approach to the analysis of human cells: the transformed amniotic cells (AMA) master database and its link to genome DNA sequence data. *Electrophoresis* **12**, 989-1071

21. Celis, J. E., Dejgaard, K., Madsen, P., Leffers, H., Gesser, B., Honoré, B., Rasmussen, H. H., Olsen, E., Lauridsen, J. B., Ratz, G., Mouritzen, S., Hellerup, M., Andersen, A., Walbum, E., Celis, A., Bauw, G., Puvpe, M., Van Damme, J., and Vandekerckhove, J. (1990) The MRC-5 human embryonal lung fibroblast two-dimensional gel cellular protein database: quantitative identification of polypeptides whose relative abundance differs between quiescent, proliferating and SV40 transformed cells. *Electrophoresis* 12, 1072-1113.
22. Garrels, J. I., Franz, B. R., Chang, C., and Latter, G. (1990) Quantitative exploration of the REF52 protein database: cluster analysis reveals the major protein expression profiles in responses to growth regulation, serum stimulation, and viral transformation. *Electrophoresis* 12, 1114-1130.
23. Van Bogelen, R. A., Hutton, M. E., and Neidhardt, F. C. (1990) Gene-protein database of *Escherichia coli* K-12. 3rd ed. *Electrophoresis* 12, 1131-1166.
24. Hewick, R. M., Hunkapiller, M. W., Hood, L. E., and Dreyer, W. J. (1981) A gas-liquid solid phase peptide and protein sequenator. *J. Biol. Chem.* 256, 7990-7997.
25. Weber, K., and Osborn, M. (1985) In *The Proteins and Sodium Dodecyl Sulfate: Molecular Weight Determination on Polyacrylamide Gels and Related Procedures* (Neurath, H. et al., eds), Vol. 1, pp. 179-223. Academic, New York.
26. Hunkapiller, M. W., Lujan, E., Ostrander, F., and Hood, L. E. (1983) Isolation of microgram quantities of proteins from polyacrylamide gels for amino acid sequence analysis. *Methods Enzymol.* 91, 227-236.
27. Vandekerckhove, J., Bauw, G., Puvpe, M., Van Damme, J., and Van Montagu, M. (1985) Protein-blotting on polybrene-coated glass-fiber sheets. *Eur. J. Biochem.* 152, 9-19.
28. Aebersold, R. H., Teplow, D. B., Hood, L. E., and Kent, S. B. H. (1986) Electrophoresis onto activated glass. *J. Biol. Chem.* 261, 4229-4238.
29. Bauw, G., De Loose, M., Inzé, D., Van Montagu, M., and Vandekerckhove, J. (1987) Alterations in the phenotype of plant cells studied by NH₂-terminal amino acid-sequence analysis of proteins electroblotted from two-dimensional gel-separated total extracts. *Proc. Natl. Acad. Sci. USA* 84, 4806-4810.
30. Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* 262, 10035-10038.
31. Eckerskorn, C., Mewes, W., Goretzki, H., and Lottspeich, F. (1985) A new siliconized-glass fiber as support for protein-chemical analysis of electroblotted proteins. *Eur. J. Biochem.* 176, 509-519.
32. Moos, M., Jr., Nguyen, N. Y., and Liu, T.Y. (1988) Reproducible high yield sequencing of proteins electrophoretically separated and transferred to an inert support. *J. Biol. Chem.* 263, 6003-6008.
33. Kennedy, T. E., Gawinowicz, M. A., Barzilai, A., Kandel, E. R., and Sweatt, J. D. (1988) Sequencing of proteins from two-dimensional gels by using in situ digestion and transfer of peptides to polyvinylidene difluoride membranes: application to protein associated with sensitization in *Aplysia*. *Proc. Natl. Acad. Sci. USA* 85, 7008-7012.
34. Aebersold, R. H., Leavitt, J., Saavedra, R. A., Hood, L. E., and Kent, S. B. H. (1987) Internal amino acid sequence analysis of protein separated by one- or two-dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. *Proc. Natl. Acad. Sci. USA* 84, 6970-6974.
35. Bauw, G., Van Den Bulcke, M., Van Damme, J., Puvpe, M., Van Montagu, M., and Vandekerckhove, J. (1988) Protein electroblotting on polybase-coated glassfiber and polyvinylidene difluoride membranes: an evaluation. *J. Prot. Chem.* 7, 194-199.
36. Celis, J. E., Ratz, G. P., Madsen, P., Gesser, B., Lauridsen, J. B., Leffers, H., Rasmussen, H. H., Nielsen, H. V., Crugier, D., Basse, B., Honoré, B., Möller, O., Celis, A., Vandekerckhove, J., Bauw, G., Van Damme, J., Puvpe, M., and Van Den Bulcke, M. (1989) Comprehensive human cellular protein databases and their implication for the study of genome organization and function. *FEBS Lett.* 244, 247-254.
37. Bauw, G., Van Damme, J., Puvpe, M., Vandekerckhove, J., Gesser, B., Lauridsen, J. B., Ratz, G. P., and Celis, J. E. (1989) Protein-electroblotting and -microsequencing strategies in generating protein databases from two-dimensional gels. *Proc. Natl. Acad. Sci. USA* 86, 7701-7705.
38. Aebersold, R., and Leavitt, J. (1990) Sequence analysis of proteins separated by polyacrylamide gel electrophoresis. Towards an integrated protein database. *Electrophoresis* 11, 517-527.
39. Bauw, G., Rasmussen, H. H., Van Den Bulcke, M., Van Damme, J., Puvpe, M., Gesser, B., Celis, J. E., and Vandekerckhove, J. (1990) Two-dimensional gel electrophoresis, protein electroblotting and microsequencing: a direct link between proteins and genes. *Electrophoresis* 11, 528-536.
40. Tempst, P., Link, A. J., Riviere, L. R., Fleming, M., and Elielone, C. (1990) Internal sequence analysis of protein separated on polyacrylamide gels at the submicrogram level: improved methods, applications and gene cloning strategies. *Electrophoresis* 11, 537-553.
41. Eckerskorn, C., and Lottspeich, F. (1990) Combination of two-dimensional gel electrophoresis with microsequence and amino acid composition analysis: improvement of speed and sensitivity in protein characterization. *Electrophoresis* 11, 554-561.
42. Rasmussen, H. H., Van Damme, J., Bauw, G., Puvpe, M., Gesser, B., Celis, J. E., and Vandekerckhove, J. (1991) In *Methods in Protein Sequence Analysis* (Jörnvall, H., and Höög, J. O., eds), pp. 103-114. Eighth International Conference on Methods in Protein Sequence Analysis. Birkhäuser Verlag, Boston.
43. Olson, A. D., and Miller, M. J. (1988) Elsie 4: quantitative computer analysis of sets of two-dimensional gel electrophoregrams. *Anal. Biochem.* 169, 49-70.
44. Vincens, P., Paris, N., Pujol, J. L., Gaboriaud, C., Rabilloud, T., Penneuer, J., Matherat, P., and Tarroux, P. (1986) HERMeS: a second generation approach to the automatic analysis of two-dimensional electrophoresis gels. Part I: Data acquisition. *Electrophoresis* 7, 347-356.
45. Celis, J. E., Madsen, P., Celis, A., Nielsen, H. V., and Gesser, B. (1987) Cyclin (PCNA, auxiliary protein of DNA polymerase-δ) is a central component of the pathway(s) leading to DNA replication and cell division. *FEBS Lett.* 220, 1-7.
46. Anderson, N. G., and Anderson, N. L. (1982) The human protein index. *Clin. Chem.* 28, 739-748.



Bo Franzén¹
 Stig Linder²
 Ken Okuzawa²
 Harabumi Kato²
 Gert Auer¹

¹Division of Tumor Pathology,
 Department of Pathology, Division
 of Experimental Oncology,
 Karolinska Hospital and Institute,
 Stockholm Sweden

²Tokyo Medical College, Department
 of Surgery, Tokyo

³Division of Experimental Oncology,
 Karolinska Hospital and Institute,
 Stockholm

Nonenzymatic extraction of cells from clinical tumor material for analysis of gene expression by two-dimensional polyacrylamide gel electrophoresis

We have compared different methods of preparation of malignant cells for two-dimensional electrophoresis (2-DE). We found all methods using fresh tissue to be superior compared to methods using frozen tissue. Our results indicate that nonenzymatic methods of preparation of tumor cells, including fine needle aspiration, scraping and squeezing, have advantages over methods using enzymatic extraction of cells. Nonenzymatic methods are rapid, appear to reduce loss of high molecular protein species, and alleviate the necessity of separating viable and nonviable cells by Percoll gradient centrifugation. Using these techniques, high-quality 2-DE maps were derived from tumors of the lung and breast. In the resulting polypeptide patterns, heat shock proteins, non-muscle tropomyosins and intermediate filament were identified. We conclude that nonenzymatic extraction of malignant cells from fresh tumor tissue improves the possibilities that these techniques may be useful in clinical diagnosis.

1 Introduction

Tumors may develop by a number of different mechanisms in any given cell type. At the time of diagnosis, tumors will have progressed along different pathways to various stages of malignancy. To provide a basis for individual therapy it is of importance to examine specific properties of the tumor cell population in each patient. A large number of different markers have been described in order to increase the diagnostic accuracy. It is likely that a combination of several markers is needed in the future in order to reflect different properties of the tumor. One important method for the resolution of a large number of potential markers is two-dimensional electrophoresis (2-DE). Extensive efforts are being made in identifying various polypeptides separated by 2-DE and to characterize how the expression of these polypeptides is affected by the response to cellular transformation and various culture conditions [1,2]. It would be of value to transfer this information to 2-DE separations of polypeptides from tumor tissue samples. However, one prerequisite is that the quality of the 2-DE gels from tumor samples is comparable in quality with 2-DE gels from samples of cultured cells.

Frozen tumor tissues are commonly used for various biochemical assessments. However, if such samples are analyzed by 2-D polyacrylamide gel electrophoresis (PAGE), the polypeptide patterns are obscured by contamination of serum- and connective tissue proteins. Such nontumor-cell-related variations represent serious problems in the interpretation and inter-patient comparison of 2-DE

patterns [3]. 2-DE patterns of cells prepared from fresh tumor material were analyzed after enzymatic extraction of tumor cells [4, 5] or after culturing tumor fragments in medium containing radioactive amino acids [6]. These procedures may, however, lead to alterations in the gene expression/polypeptide patterns. We are only aware of one study where nonenzymatic extraction of cells from fresh tumor tissue (prostate cancer) was used to prepare samples for 2-D PAGE [4]. We have examined enzymatic extraction and various nonenzymatic preparation techniques, including fine needle aspiration, for the preparation of cells from fresh tumor tissues. We describe nonenzymatic extraction procedures that are rapid, lead to high-quality 2-DE patterns, and that alleviate the necessity to purify tumor cell populations from dead cells.

2 Materials and methods

2.1 Cell cultures and samples used for spot identification

A rat embryonal fibroblast cell line, WT2 (a kind gift from Dr. J. I. Garrels and Dr. S. Patterson) was used for the identification of a number of heat shock and structural proteins. Human normal diploid lung fibroblasts, WI38, human epithelial breast carcinoma cells, MDA-231 and MCF-7 were purchased from ATCC and grown as recommended. Polypeptides prepared from a leukemia type pre-B-ALL were separated by 2-DE. The 2-DE map was then analyzed by Dr. S. M. Hanash (University of Michigan, Ann Arbor, USA).

2.2 Tumor tissues samples

In this study, 2-DE maps from seven tumors were used as representative illustrations: two adenocarcinoma of the lung (LA, and LB, mucinous, both cases intermediate grade of differentiation), one squamous carcinoma of the lung (LS), one carcinoid-like breast cancer (BC), one microfollicular adenoma (highly differentiated) of the thyroid (TA), one highly differentiated hyperneph-

Correspondence: Dr. Bo Franzén, Division of Tumor Pathology, Department of Pathology, L1:01, Karolinska Hospital and Institute, 10401 Stockholm 60, Sweden

Abbreviations: 2-DE, Two-dimensional polyacrylamide gel electrophoresis; IEF, isoelectric focusing; LDH, lactate dehydrogenase; NP-40, Nonidet P-40; PBS, phosphate buffered saline; PCNA, proliferating cell nuclear antigen; PIH, protease inhibitors; PMSF, phenylmethyl sulfonyl fluoride; SDS, sodium dodecyl sulfate; WW, wet weight

roma, a tumor of the kidney (KH), and finally one case of poorly differentiated corpus carcinoma (CP).

2.3 Preparation of cultured cells

The cell monolayers were washed twice in phosphate buffered saline (PBS) and then scraped off in ice-cold PBS including protease inhibitors (PIH). phenylmethyl-sulfonyl fluoride (PMSF) 0.2 mM and 0.83 mM benzamidine pelleted at 660 × g, 3 min (+4°C) and washed one time before final centrifugation at 2700 × g, 5 min. The wet weight of the cell pellet was recorded and the cells were stored at -80°C until further processing.

2.4 Preparation of tumor tissue samples

2.4.1 General remarks

Macroscopically representative and non-necrotic tumor tissues were selected within 20 min after resection. Parallel samples were routinely prepared for cytology. The samples were processed as rapidly as possible on ice or at +4°C and in the presence of PIH. Cells were stained with DiffQuick (Baxter) and usually examined at three different occasions during the preparation procedure: (i) cytology sample, (ii) extracted cells and (iii) cells after percoll gradient centrifugation.

2.4.2 Specimen acquisition

The strategy of sample preparation is shown in Fig. 1. Tumor tissue cell samples were usually obtained by fine needle aspiration (NA) using a 0.7 mm needle. The syringe was filled with 1–2 mL of ice-cold culture medium/PIH. We found that if a tumor appeared to be very fibrous it is difficult to extract enough cells for 2-DE analysis. In these cases, two alternative techniques were examined. (i) The tumor was cut in the middle and the fresh surface scraped (SC) by a scalpel. The cell-rich material was then transferred to ice-cold culture medium (L15 with 5% fetal calf serum)/PIH. (ii) A part of the tumor sample was placed in culture medium on ice for further processing at the laboratory in the following way: the material was cut into very small fragments on a pre-cooled dissection plate and transferred to a small glass chamber with a 0.7 mm metal net 5 mm above the bottom of the chamber. Medium /PIH was added to cover the sample (8 mL) which was gently squeezed (SQ) towards the net in order to release and wash out cells. NA and SC were also compared with an enzymatic extraction (EE) procedure described previously [5]. Briefly, thin slices of tissue were incubated with collagenase (1 mg/mL) and elastase (2 mg/mL) in medium for 1 h at 37°C. Extracted cells from every sample were then subjected to percoll gradient centrifugation (Section 3.2.3).

2.4.3 Separation of cells by Percoll gradient centrifugation

The cell suspension was filtered through two nylon mesh filters, (i) 250 µm and (ii) 100 µm and then centrifuged

at 660 × g for 3 min. The cell pellet was resuspended carefully in medium, using a syringe and loaded onto a two-step discontinuous Percoll/PBS gradient, 20.4 (density = 1.03 g/mL) and 54.7% (density = 1.07 g/mL), and centrifuged at 1000 × g for 15 min. In this system, dead cells stay on the top, viable cells sediment to the interphase and erythrocytes sediment to the bottom. The viability of cells in the top fraction and interphase was checked by the trypan blue exclusion test. The interphase cell layer (> 90% viability) was collected and washed one time in a large volume PBS/PIH (centrifuged at 800 × g for 3 min). Finally, the cells were resuspended in 1.4 mL PBS and pelleted at 2700 × g for 5 min. The wet weight (WW) was recorded and the pellet was then stored at -80°C.

2.4.4 Final preparation of cells for 2-D PAGE analysis

From this point, cultured cell samples were treated in the same way as tumor cell samples: Each cell pellet was thawed on ice and resuspended in 1.89 µL mQ water per mg WW (= 1.89 × WW) µL. The suspension was frozen and thawed 4–5 × to break the cells [7]. A volume of (0.089 × WW) µL 10% sodium dodecyl sulfate (SDS), including 33.3% mercaptoethanol, was mixed with the sample and incubated 5 min on ice with (0.329 × WW) µL of a solution of DNase I (0.144 mg/mL 20 mM Tris-HCl with 2 mM CACl₂, 2H₂O, pH 8.8) and RNase A (0.0718 mg/mL Tris) [8,9]. The sample was frozen and lyophilized. Sample buffer [10] including

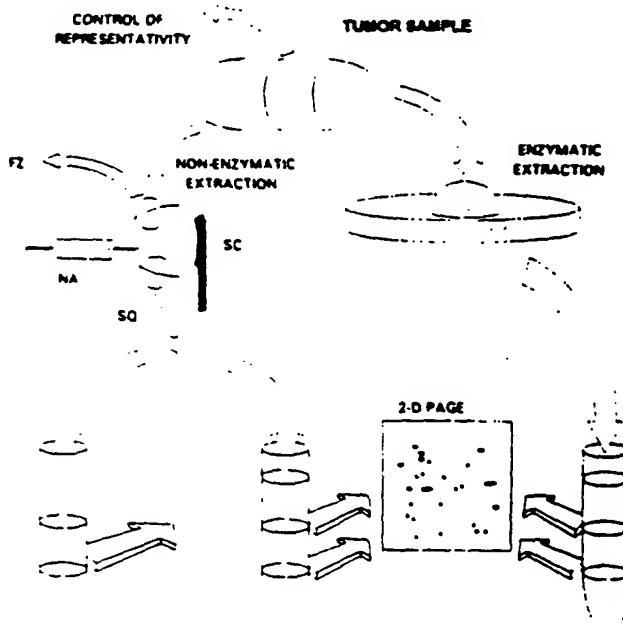


Figure 1. Experimental flow chart showing main steps of the preparation procedures. The abbreviations used for nonenzymatic extraction procedures are: FZ, frozen sample preparation; NA, needle aspiration; SC, scraped; and SQ, squeezed sample. Extracted cells are then loaded as a suspension (top volume of each tube) onto either 1.07 g/mL Percoll (left), or a discontinuous Percoll gradient from the nonenzymatic extraction (middle), or from enzymatic extraction (right). Cellular top- and interphase fractions are then used for 2-DE. For details see Section 2.

PMSF (0.2 mM), EDTA (1.0 mM), 0.5% Nonidet P-40 (NP-40), and 3-[3-cholamido propyl]-dimethylammonio]-1-propane sulfonate (CHAPS; 25 mM) was added carefully, mixed for 2.5 h and centrifuged for 15 min at

10000 rpm to remove any insoluble material. Duplicate or triplicate samples were taken for protein determination [11]. Samples were stored at -80°C prior to isoelectric focusing (IEF).

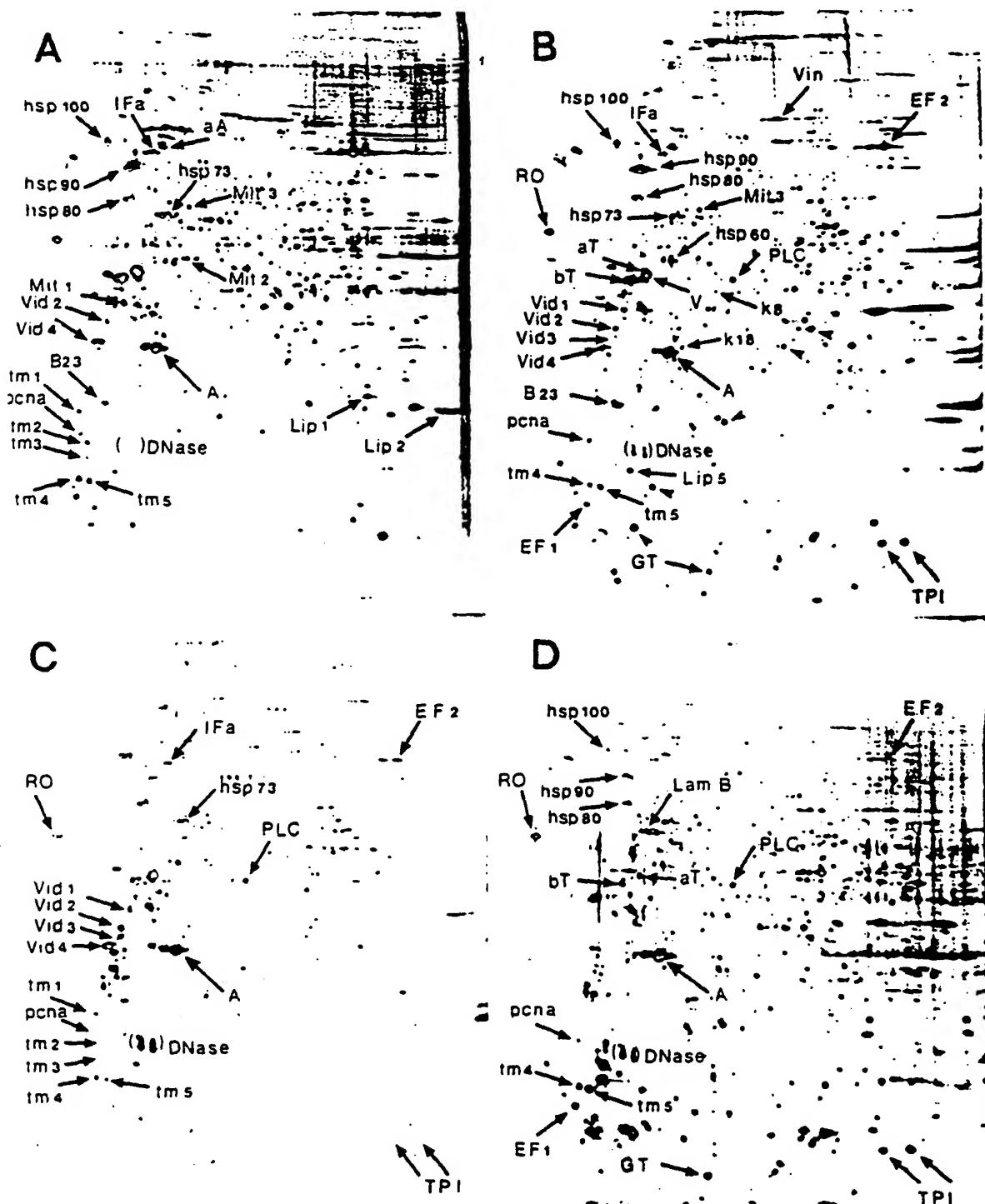


Figure 2. 2-DE analysis of samples from three cell lines and one leukemia used for the identification of polypeptides: (A) WT2; (B) MDA-231, arrowheads mark some low molecular weight cytosolic polypeptides; (C) WI38 and (D) pre B-All. The abbreviations for identified spots are explained in Table 1.

2.4.5 Preparation of frozen tumor tissue

The technique has been described previously [3,12]. Briefly, the sample is mortared frozen to a fine powder, homogenized, lyophilized and solubilized in sample buffer.

2.4.6 Control of representativity

The tumors were examined routinely by experienced pathologists and smears or imprints from the samples were also assessed for cytometric DNA content by microspectrophotometry.

2.5 2-D PAGE

2-D PAGE was performed as described [8,10] except for the following details. The glass tubes for IEF, 1.2 × 200 mm, contained 2.0% Resolyte, pH 4–8 (BDH) and were cast to a height of 180 mm. A stock solution of acrylamide (Serva) and N,N'-methylenebisacrylamide (16.7:1 for IEF and 37.5:1 for the second dimension) was deionized by mixing with 5% w/v Duolite MB 5313 mixed-resin ion exchanger (BDH) for 30 min, filtered (with a 0.22 µm nitrocellulose filter) and stored at -70°C. N,N'-Methylenebisacrylamide, N,N,N',N'-tetramethylethylenediamine (TEMED) and ammonium persulfate were purchased from Bio-Rad. IEF tubes were prefocused at 200 V in 60 min. To each tube a sample corresponding to 20–40 µg protein was applied and focused for 14.5 h at 800 V and finally 1.0 h at 1000 V using a Protean II cell (Bio-Rad) and Model 1000/500 Power Supply (Bio-Rad). The tube gels were finally extruded into 1.25 mL equilibration buffer, containing 60 mM Tris, pH 6.8 (2% SDS, 100 mM dithiothreitol and 10% glycerol), frozen on dry ice and stored at -70°C. The second dimension (1.0 × 180 × 90 mm) of the acrylamide concentration was 10%

T, and the gel contained 376 mM Tris, pH 8.8, and 0.1% SDS. IEF gels were applied on top of the slab gel, sealed with 0.5% agarose containing electrophoresis running buffer (60 mM Tris-base, 0.2 M glycine and 0.1% SDS) and electrophoresed with 10–11 mA per gel (constant current) at +10°C. Six gels were run together in a Protean II xi 2-D Multi-Cell (Bio-Rad). Proteins were visualized by silver staining and photographed with the acidic side to the left [13,14].

2.6 Identification of polypeptides

Vimentin and vimentin-derived polypeptides were identified by extraction of an MDA-231 cell lysate with 0.6 M KCl/0.5% NP-40 [15]. Tropomyosins were extracted from MDA-231 and WI38 cell lysates [16], and cytokeratins were extracted from MDA-231 and MCF-7 cell lysates [17]. The patterns were compared with published maps [19–21]. Proliferating cell nuclear antigen (PCNA) was identified by immunoblotting (PC10 mAB, Dakopatt) using a semidry system (Multiphor II Nova Biol. Pharmacia-LKB Biotechnology AB) and enhanced chemiluminescence (ECL) detection (Amersham).

3 Results

3.1 2-DE of samples prepared from normal and tumorigenic cultured cells

The object of this study was to develop methods for preparation of 2-DE maps from human tumor tissue which have the same high resolution as those obtained from cultured cells. Shown in Fig. 2 are high resolution 2-DE gels prepared from cultured cells and one leukemia: SV40 transformed embryonal rat fibroblasts WT2 (Fig. 2a); human MDA-231 breast carcinoma cells (Fig. 2b); human WI38 fibroblasts (Fig. 2c) and human pre B-ALL

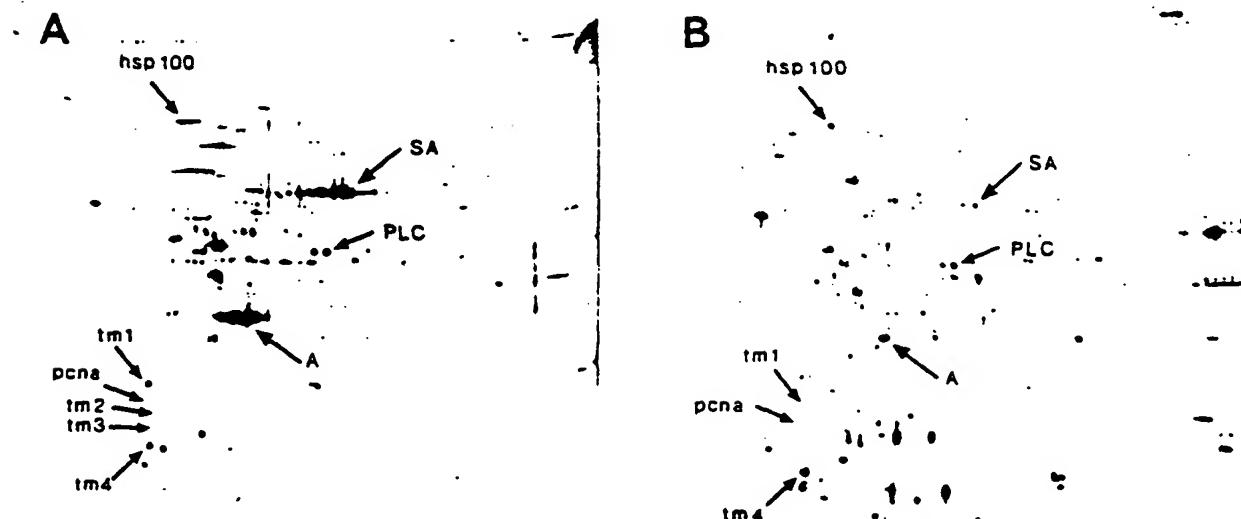


Figure 3. 2-DE analysis of a case of lung adenocarcinoma (LA). Comparison of 2-DE gel quality between (A) frozen and (B) fresh (needle aspiration) tissue preparation.

cells (Fig. 2d). Polypeptides were identified through a laboratory exchange of cell samples/2-DE maps and through 2-DE analysis of purified proteins (Table 1).

3.2 Preparation of samples from solid tumors

3.2.1 Fresh versus frozen tissue

An adenocarcinoma of the lung (LA) was prepared for 2-DE by conventional methods using frozen material (Fig. 3a). There are several possibilities for the poor resolution using frozen tissue, including the presence of high molecular weight protein aggregates. Filtering extracts through 0.1 µm filters (Durapore, Millipore) resulted in a slightly improved resolution (not shown). When fresh tumor tissue from tumor LA was used for sample preparation, using fine needle aspiration to collect the cells, the resolution was considerably improved (Fig. 3b). The use of fresh tissue resulted in a general increase in resolution, which was most pronounced in the 50–100 kDa molecular mass range. A number of differences in the protein profiles of the gels in Figs. 3a and 3b can be observed, some of which are indicated in the figures. The decrease in serum albumin in Fig. 3b is likely to result from loss of serum proteins occurring when cells were pelleted after aspiration. Other differences, such as the decreased level of transformation-sensitive tropomyosins (TM1–TM3), may result from enrichment of tumor cells in the sample of Fig. 3b. Fine needle aspiration, a well-established technique in cytology, extracts mainly tumor cells because of decreased intercellular adhesiveness of neoplastic cells as compared to normal tissue. Microscopic examination of Diff-Quick-stained extracted cells from case LA revealed almost 100% tumor cells, whereas the whole tissue extract contained approximately 60% tumor cells.

Table 1. Names and abbreviations for identified spots

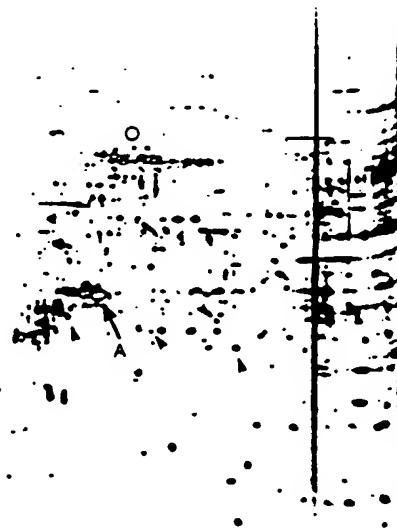
Spot	Name	Basis for identification
A	Actins	a
aA	alpha-Actinin	a
B23	Protein B23 /Numatrin	a
EF2	Elongation factor 2	a
EF1	Elongation factor 1 β	a
GT	Glutathione-S-transferase (pI)	a
hsp60	Heat shock protein 60	a
hsp73	Heat shock protein 73	a
hsp80	Heat shock protein 80, GRP78, BIP	a
hsp90	Heat shock protein 90	a
hsp100	Heat shock protein 100, Endoplasmic	a
IFa	Intermediate filament associated	a
k8	Cytokeratin 8	b and a
LamB	Lamin B	a
Lip1	Lipocortin I	a
Lip2	Lipocortin II	a
Lip5	Lipocortin V	a
Mit1	Mitcon 1/β – F1 ATPase	a
Mit2	Mitcon 2	a
Mit3	Mitcon 3	a
MRP	Mucine Related Polypeptides	—
pena	Proliferating cell nuclear antigen	c and a
PLC	Phospholipase C (1)	a
RO	RO/SS-A antigen	a
SA	Serum Albumin	b and a
aT	alpha-Tubulin	a
bT	beta-Tubulin	a
tm1	Non-muscle tropomyosin isoform 1	b and a
tm2	Non-muscle tropomyosin isoform 2	b and a
tm3	Non-muscle tropomyosin isoform 3	b and a
tm4	Non-muscle tropomyosin isoform 4	b and a
tm5	Non-muscle tropomyosin isoform 5	b and a
TPI	Triose phosphate isomerase	a
V	Vimentin	b and a
Vid1	Vimentin derived protein	b and a
Vid2	Vimentin derived protein	b and a
Vid3	Vimentin derived protein	b and a
Vid4	Vimentin derived protein	b and a
Vin	Vinculin	a

a. homologous position with respect to other mammalian systems

b. purified protein(s)

c. immunoblotting

A



B



Figure 4. 2-DE analysis of a case of breast carcinoma (BC). Comparison of 2-DE quality and some differences in detected spots (arrow heads indicate increased intensity and circles or bracket indicate decreased intensity of the same spots) between (A) enzymatically and (B) nonenzymatically (scraped) tissue preparation

3.2.2 Comparison of different methods for preparing cells from fresh tumor tissue

Samples were prepared from breast and lung carcinomas using either an enzymatic treatment with collagenase/elastase or using nonenzymatic preparations (Fig. 4). A number of differences in the protein profiles were observed in the resulting 2-DE gels, some of which are indicated in Figs. 4a and b. These differences include both increases and decreases in spot intensity. These differences may result from degradation of high molecular weight polypeptides during enzymatic treatment, increased solubilization of polypeptides, or may have other causes. For many tumors, it was only possible to obtain

small amounts of material since they were reserved for other examinations. In these cases, samples could be prepared for 2-DE using either needle aspiration or scraping. Figure 5a shows a 2-DE gel prepared from squamous lung carcinoma (LS) cells collected by needle aspiration and Fig. 5b shows a gel prepared from the same tumor by scraping. In this case, a number of differences were recorded between the two procedures, some of which are arrowed in Fig. 5. Samples obtained from other tumors (breast and lung) generally showed fewer differences between these two methods of cell sampling (not shown). These data show that different nonenzymatic extraction procedures may yield different polypeptide patterns. However, the number of spots with a large

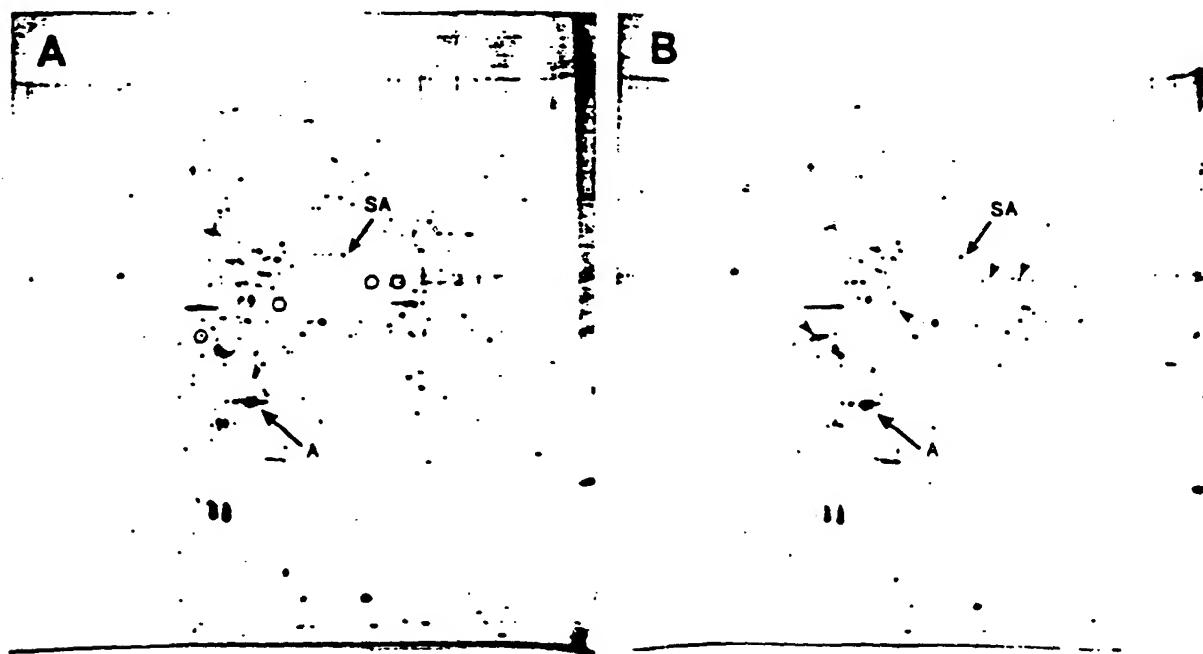


Figure 5. 2-DE analysis of a case of lung cancer (LS). Comparison of 2-DE gel quality and detected spots (arrow heads and circles) between (A) aspirated (needle aspiration) and (B) scraped preparations from fresh tissue.

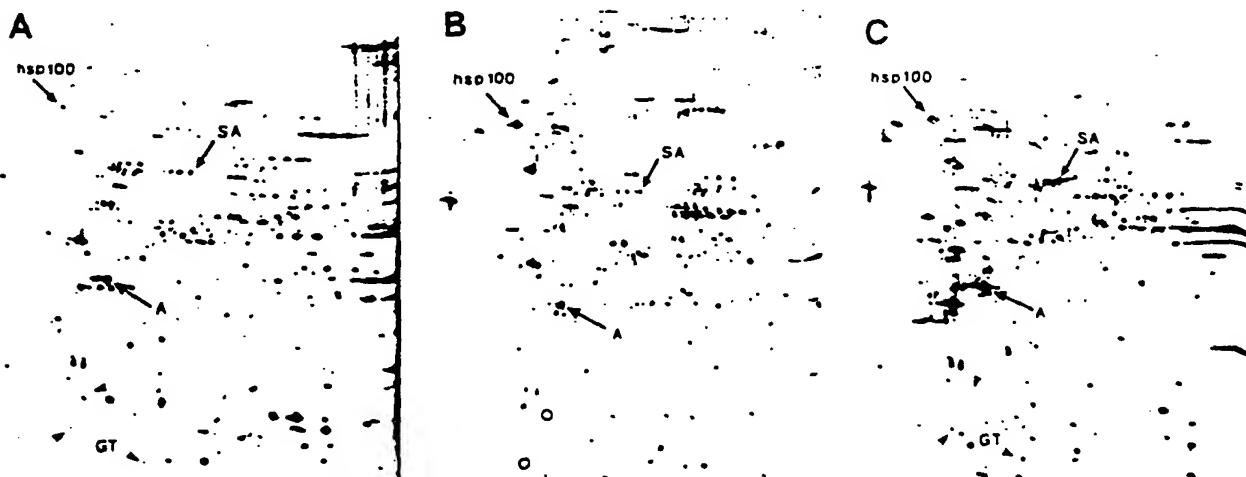


Figure 6. 2-DE analysis of three other types of tumors. (A) hypernephroma, (B) an adenoma of the thyroid and (C) corpus cancer, using the nonenzymatic preparation technique. Arrowheads and circles indicate some cytosolic polypeptides.

difference in intensity were lower than when a nonenzymatic preparation was compared with an enzymatic preparation.

2-DE maps of satisfactory quality were prepared by a third procedure. Cells were released from small pieces of tumor by squeezing (see Section 2). Some examples of this are shown in Fig. 6 where 2-DE maps derived from a case of hypernephroma, KH (Fig. 6a), a case of thyroid tumor, TA (Fig. 6b) and a case of corpus cancer, CP (Fig. 6c) can be seen. We conclude that nonenzymatic techniques are useful for 2-DE analysis of a number of different tumors. The quality of the resulting gels is com-

parable to that obtained using cultured cells (compare the gels in Fig. 2 with those in Fig. 4, 6 and 7). Which of these methods will be optimal will, in our experience, depend on the tumor material. For example, very small tumors are preferably extracted by squeezing; on the other hand, breast cancers (which are often fibrous) yield satisfactory samples using scraping.

3.2.3 Purification of cells on percoll gradients

We considered the possible advantage of separating viable cells from dead cells, erythrocytes, and debris using discontinuous Percoll gradients. Cells collected

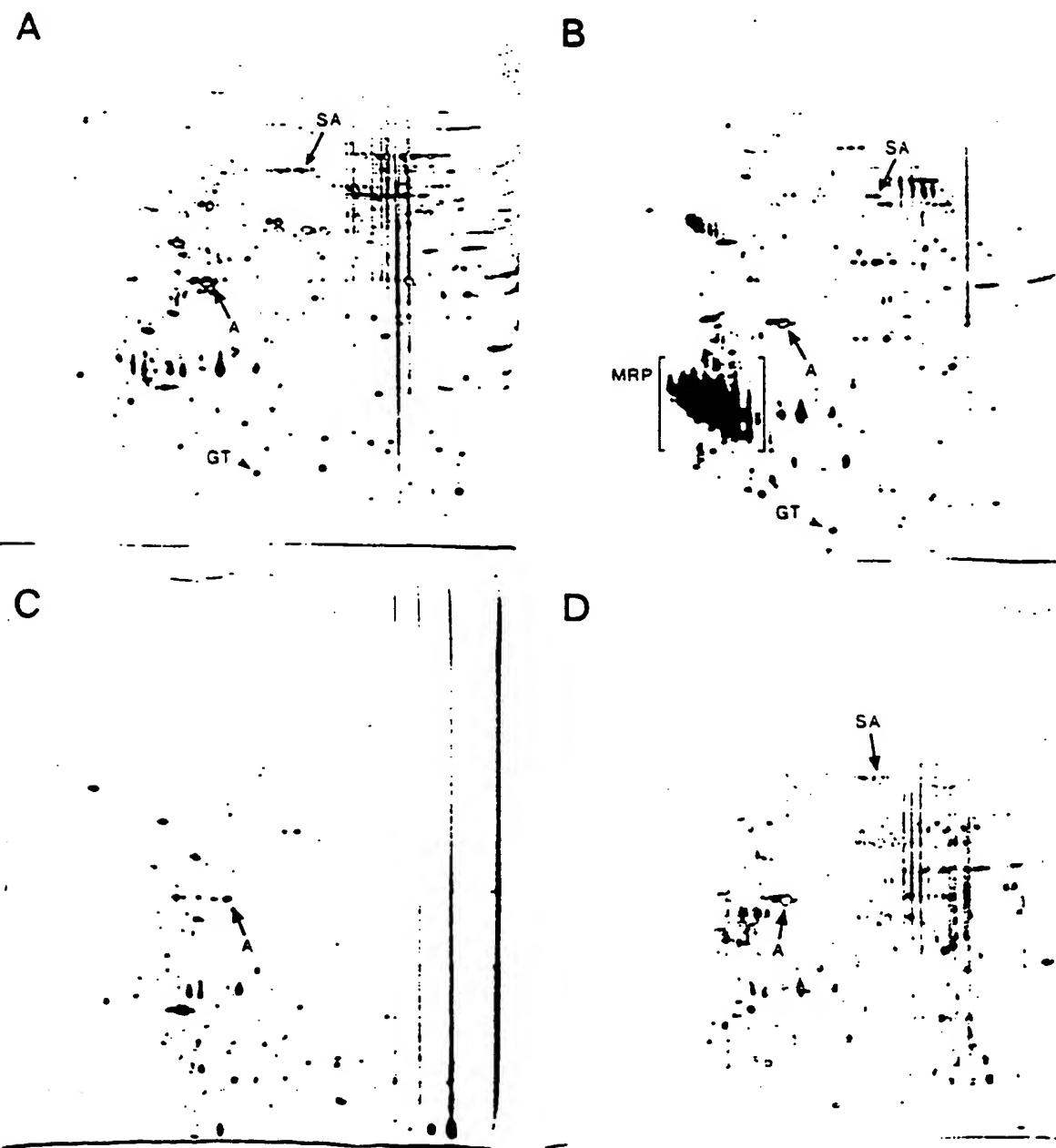


Figure 7. 2-DE analysis of polypeptides from viable (b and d) and nonviable (a and c) cells of an adenocarcinoma of the lung (LB), separated using discontinuous Percoll density gradient. Nonenzymatic preparation technique (a and b) and enzymatic preparation technique (c and d) are compared.

from the interphase showed a viability of more than 90% as judged by trypan blue exclusion test. However, it was found that the yield of viable cells decreased dramatically if the tissue resection was not immediately processed. To study the effect of lysis of cells during the preparation procedure, 2-DE maps were prepared from nonenzymatically extracted cells of case LB collected from the top fraction (nonviable, Fig. 7a) and interphase fraction (viable, Fig. 7b). These 2-DE maps were compared with corresponding fractions (nonviable, Fig. 7c, and viable, Fig. 7d) of enzymatically extracted cells. One clear disadvantage of the enzymatic technique was that when loss of cell viability occurred during preparation, a dramatic loss of high molecular weight polypeptides was observed (Fig. 7c). This was probably due to degradation of intracellular proteins. However, nonenzymatic preparations showed fewer differences between viable and nonviable cells: The most pronounced alteration was a decrease of a group of mucine related proteins (Fig. 7b). We conclude, therefore, that discontinuous Percoll gradient is necessary after enzymatic extraction of cells, but can be omitted from the nonenzymatic tumor sample preparation procedure.

We used the MDA-231 cell line to study the effects of cell lysis and leakage of cytosolic polypeptides during sample preparation. Remarkably, after 30, 50, 80 and 140 min of incubation in PBS/PIH at 0°C, no significant changes were observed in the 2-DE pattern (not shown). Although loss of cell viability may not result in protein degradation when cells are incubated in the presence of protease inhibitors, loss of cytosolic proteins would be expected during pelleting of cells. We monitored the loss of lactate dehydrogenase (LDH) activity into the supernatant during incubation in PBS of MDA-231 and MCF-7 breast cancer cells at 20°C. In both cases, loss of viability was paralleled by release of LDH from the cells (Fig. 8). After 5 h, 70% of the MCF-7 cells, but only 30% of the MDA-231 cells were dead (not shown).

These data indicate the impact of a rapid preparation procedure, at low temperature, of fresh tumor samples. Experiments have also been performed using only 1.07 g/mL Percoll (Fig. 6c and Fig. 1, left test tube) in order to remove erythrocytes. One clear advantage with this procedure, which today is routinely utilized, is a higher yield of viable cells, probably due to decreased sample preparation time.

4 Discussion

We describe procedures for sample preparation from solid tumors for 2-DE. 2-DE maps could be derived from solid tumors which were similar in quality to those obtained from cultured cells. Compared to methods using frozen material, the resolving power of the 2-DE technique is increased, allowing examination of a large number of polypeptides from tumors of different malignancies. Other investigators [12,22] have used samples from frozen tumors to derive 2-DE maps. We have previously described disadvantages encountered using frozen tumor samples including variations in contaminating proteins between different samples [3]. The methods described here are based on the preparation of cells from tumors without enzymatic digestion. The enzymatic step could be avoided since malignant cells usually grow as solid masses which are not strongly attached to the matrix. Furthermore, we found that omitting the enzymatic digestion alleviated the necessity of purifying viable tumor cells on Percoll gradients. This was in sharp contrast to enzymatically treated samples, where loss of viability leads to loss of high molecular weight proteins (Fig. 7c).

At least in the case of lung cancer, viable and nonviable cells showed small differences in respect to 2-DE maps. Presumably, protease inhibitors penetrate cells and inhibit proteolysis. In model experiments, we observed leakage of cytosolic protein (LDH) from the cells in parallel to loss of viability. Apparently, however, only a limited decrease of the level of low molecular weight cytosolic polypeptides was detected using silver staining combined with visual inspection. We have found that although some tumors are well suited for the preparation procedure described, others are not. In general, good results were obtained using tumors of the lung, breast, corpus and lymphomas. In contrast, cells from thyroid adenomas and hypernephroma showed poor viability. We were in these cases unable to separate nonviable cells from viable cells, and we can therefore not evaluate the consequence of the loss of viability on 2-DE patterns, apart from a loss of some low molecular weight cytosolic polypeptides.

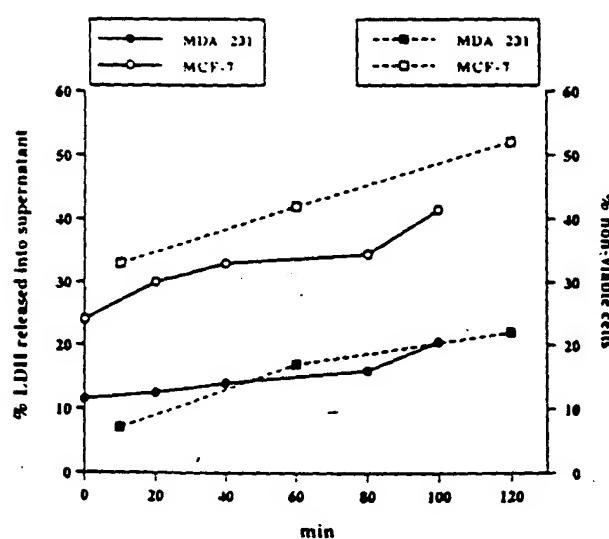


Figure 8. The relative release (fraction in supernatant of total) of lactate dehydrogenase activity (LDH) and cell viability versus incubation time of the mammary carcinoma cell lines MDA-231 and MCF-7 during incubation in PBS at 20°C.

Highly differentiated tumors may show lower viability as compared with poorly differentiated tumors (Dr. Farkas Vánky, personal communication). A number of samples from thyroid tumors were prepared for 2-DE but most cases showed poor viability. We believe that special care is needed during preparation of generally highly differentiated tumor groups. The difference between loss of viability/leakage of LDH of the more differentiated MCF-7 cells and the less differentiated MDA-231 cells is in line

with these observations (Fig. 8). A number of potential and interesting markers, like tropomyosin isoforms, cyto-keratins and heat shock proteins, appear to be insensitive to loss of viability during the preparation procedure. We have to date made numerous observations of alterations in the expression of these polypeptides in breast cancers and lung cancers.

Another problem that may occur, irrespective of sample preparation techniques used, is admixture of lymphocytes. These cases are easily detectable in smears and it may therefore be possible to select lymphocyte specific spots as "internal markers" for the 2-D PAGE analysis. Studies using this approach are in progress. Many of the polypeptides identified are structural (Table 1). Since the expression of many of these polypeptides are known to vary between normal and malignant cells, the possibility to determine their expression simultaneously is appealing. In the specific case of breast cancer, alterations in the expression of intermediate filament proteins (cytokeratins) are known to occur during tumor progression [23]. Other proteins known to be differentially expressed between normal cells and transformed cells are tropomyosins, numatrin/B23, heat shock proteins and PCNA. To this end, we have observed alterations in the expression of cytokeratin 8, hsp 90, and non-muscle tropomyosin isoform 2 during malignant progression. (Okuzawa *et al.*, in preparation and Franzén *et al.*, in preparation).

The method of choice for sample preparation from tumor tissues will depend on the properties of the tumor material studied. It may be important to use only one method when comparing cases within one group, as differences were observed between methods. The advantages of the nonenzymatic techniques are (i) that it minimizes contamination with connective tissue, (ii) that problems with contamination of serum proteins are avoided, and (iii) that separation of viable and dead cells is not necessary. Hereby the revolving power of 2-D PAGE is maximized for the analysis of human tumors and studies on inter-tumor variations in gene expression are facilitated. In addition, the polypeptide patterns obtained may be more representative for the *in vivo* tumor cell since the use of enzymes and incubations have been minimized.

We would like to thank Dr. J. I. Garrels, Dr. S. Patterson, Dr. S. M. Hanash and Dr. J. E. Celis for making sample and 2-DE map exchanges possible. This study was sup-

ported by grants from the Swedish Cancer Society and the Cancer Society in Stockholm.

Received March 5, 1993

5 References

- [1] Celis, J. E., Dejgaard, K., Madsen, P., Leffers, H., Gesser, B., Honore, B., Rasmussen, H. H., Olsen, E., Lauridsen, J. B. and Ratz, G., *Electrophoresis* 1990, 11, 1072-1113.
- [2] Garrels, J. I., Franzén, B. R., Chang, C., Latter, G., *Electrophoresis* 1990, 11, 1114-1130.
- [3] Franzén, B., Iwabuchi, H., Kato, H., Lindholm, J. and Auer, G., *Electrophoresis* 1991, 12, 509-515.
- [4] Sherwood, E. R., Berg, L. A., Mitchell, N. J., McNeal, J. E., Kozlowski, J. M. and Lee, C., *J. Urology* 1990, 143, 167-171.
- [5] Endler, A. T., Young, D. S., Wold, L. E., Lieber, M. M. and Curie, R. M., *J. Clin. Chem. Clin. Biochem.* 1986, 24, 981-992.
- [6] Forchhammer, J. and Macdonald-Bravo, H., in: Celis, J. E. and Bravo, R. (eds.), *Gene Expression in Normal and Transformed Cells*, Plenum, New York 1983, pp. 291-314.
- [7] Linder, S., Brzeski, H. and Ringertz, N. R., *Exp. Cell. Res.* 1979, 120, 1-14.
- [8] Celis, J. E. and Bravo, R. (Eds.), *Two-dimensional Gel Electrophoresis of Proteins*, Academic Press, New York 1984, pp. 3-36.
- [9] Garrels, J. I., *J. Biol. Chem.* 1979, 254, 7961-7977.
- [10] Anderson, N. L., *Two-Dimensional Electrophoresis. Operation of the ISO-DALT System*, Large Scale Biology Press, Washington, DC 1988, 162.
- [11] Bradford, M., *Anal. Biochem.* 1976, 72, 248.
- [12] Tracy, R. P., Wold, L. E., Currie, L. M. and Young, D. S., *Clin. Chem.* 1982, 28, 890-899.
- [13] Merrill, C. R., Goldman, D., Sedman, S. A. and Elbert, H. M., *Science* 1981, 211, 1437-1438.
- [14] Morrissey, J. H., *Anal. Biochem.* 1981, 117, 307-310.
- [15] Gard, D. L., Bell, P. B., Lazarides, E., *Proc. Natl. Acad. Sci. USA*, 1979, 76, 3894-3898.
- [16] Matsumura, F., Lin, J.-C., Yamashiko-Matsumura, S., Thomas, G. P. and Topp, W. C., *J. Biol. Chem.*, 1983, 258, 13954-13960.
- [17] Paulin, D., Forest, N. and Perreau, J., *J. Mol. Biol.* 1980, 144, 95-101.
- [18] Blobel, G. A., Moll, R., Franke, W. W., Kayser, K. W. and Gould, V. E., *Am. J. Pathol.* 1985, 121, 235-247.
- [19] Ochs, D. C., McConkey, H. E. and Guard, N. L., *Exp. Cell. Res.* 1981, 135, 355-362.
- [20] Bhattacharya, B., Gaddamanuga, L. P., Valverius, E. M., Salomon, D. S. and H. L. Cooper, *Cancer Res.* 1990, 50, 2105-2112.
- [21] Sommers, C. L., Walker-Jones, D., Heckford, S. E., Worland, P., Valverius, A., Clark, R., McCornick, F., Stampfer, M., Abularch, S. and Gelmann, E. P., *Cancer Res.* 1989, 49, 4258-4263.
- [22] Trask, D. K., Bond, V., Zajchowski, D. A., Yaswen, P., Suh, T. and Sager, R., *Proc. Natl. Acad. Sci. USA* 1990, 87, 2319-2323.
- [23] Trask, D. K., Bond, V., Zajchowski, D. A., Yaswen, P., Suh, T. and Sager, R., *Proc. Natl. Acad. Sci. USA* 1990, 87, 2319-2323.



Bengt Bjellqvist*
 Bodil Basse
 Eyðfinnur Olsen
 Julio E. Celis

Institute of Medical Biochemistry
 and Danish Centre for Human
 Genome Research, Aarhus
 University, Aarhus

Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions

A highly reproducible, commercial and nonlinear, wide-range immobilized pH gradient (IPG) was used to generate two-dimensional (2-D) gel maps of [³⁵S]methionine-labeled proteins from noncultured, unfractionated normal human epidermal keratinocytes. Forty one proteins, common to most human cell types and recorded in the human keratinocyte 2-D gel protein database were identified in the 2-D gel maps and their isoelectric points (*pI*) were determined using narrow-range IPGs. The latter established a pH scale that allowed comparisons between 2-D gel maps generated either with other IPGs in the first dimension or with different human protein samples. Of the 41 proteins identified, a subset of 18 was defined as suitable to evaluate the correlation between calculated and experimental *pI* values for polypeptides with known composition. The variance calculated for the discrepancies between calculated and experimental *pI* values for these proteins was 0.001 pH units. Comparison of the values by the *t*-test for dependent samples (paired test) gave a *p*-level of 0.49, indicating that there is no significant difference between the calculated and experimental *pI* values. The precision of the calculated values depended on the buffer capacity of the proteins, and on average, it improved with increased buffer capacity. As shown here, the widely available information on protein sequences cannot, *a priori*, be assumed to be sufficient for calculating *pI* values because post-translational modifications, in particular *N*-terminal blockage, pose a major problem. Of the 36 proteins analyzed in this study, 18–20 were found to be *N*-terminally blocked and of these only 6 were indicated as such in databases. The probability of *N*-terminal blockage depended on the nature of the *N*-terminal group. Twenty six of the proteins had either M, S or A as *N*-terminal amino acids and of these 17–19 were blocked. Only 1 in 10 proteins containing other *N*-terminal groups were blocked.

1 Introduction

As compared with carrier ampholyte isoelectric focusing (CA-IEF), the application of immobilized pH gradients (IPGs) in the first dimension in 2-D gel electrophoresis offers improved reproducibility [1] because the nature of the pH gradient makes the resulting focusing positions insensitive to the focusing time [2] and to the type of sample applied [3]. The recently introduced ready-made IPG strips [4] seem to be an ideal substitute for the carrier ampholyte gradients, which until now have been the most commonly used first dimensions in 2-D gel electrophoresis. The availability of standardized first dimensions opens the possibility of comparing 2-D gel maps of various cell types generated in different laboratories, provided that the focusing positions of a number of easily recognizable polypeptide spots common to the cell types

in question are known. Even though this approach is limited to experiments performed with the same standardized IPG, the flexibility provided by IPGs allows the pH gradient to be adjusted to the requirements of a particular experiment.

Exchange and communication of 2-D gel protein data requires a pH scale that is independent of the particular IPG used and by which the results can be described. The introduction of carbamylation trains and the relation of focusing positions to the spots in these trains represented a step forward towards solving the reproducibility problem experienced with carrier ampholyte focusing [5]. Problems associated with the use of carbamylation trains were mainly due to lack of temperature control and to the use of nonequilibrium focusing conditions. Accordingly, the pattern variation involved not only the resulting pH gradients, but also the relative spot positions as related to each other and to spots in the carbamylation trains. Even though the question of reproducibility has, to a large extent, been solved, the carbamylation trains are still not ideal as markers because the spots in the trains do not represent defined entities but rather a large number of differently carbamylated peptides having close *pI* values. As a result, the spots are large and poorly defined as compared to the ordinary polypeptide spots in 2-D gel maps.

Correspondence: Professor J. E. Celis, Institute of Medical Biochemistry and Danish Centre for Human Genome Research, Aarhus University, DK-8000 Aarhus C, Denmark

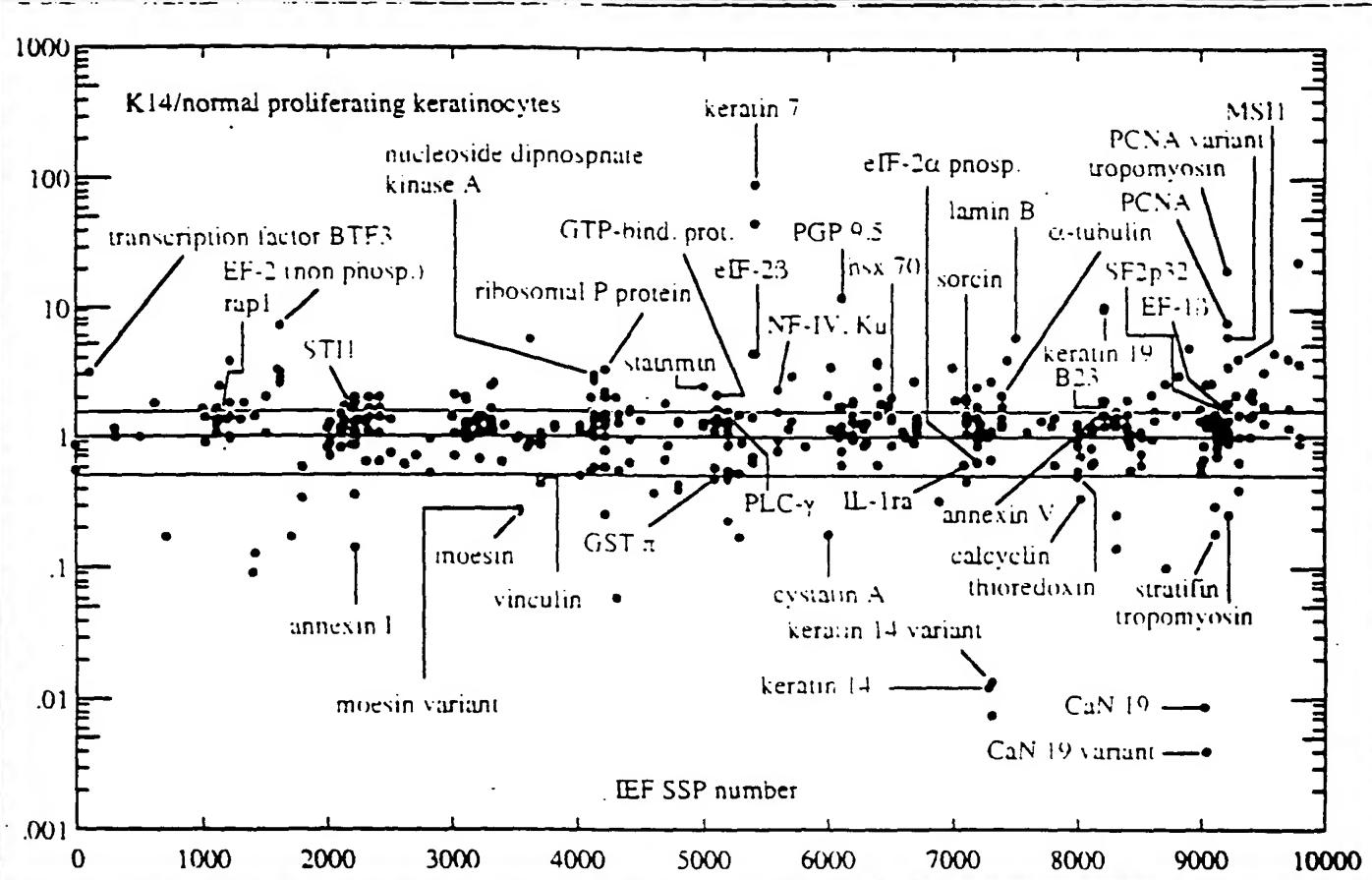
Abbreviations: CA-IEF, carrier ampholyte-isoelectric focusing; SSP, sample spot number

* Present address: Pharmacia Biotech AB, S-751 82 Uppsala, Sweden

ELECTROPHORESIS

An International Journal

3-4'94



PAPER SYMPOSIUM

ELECTROPHORESIS IN CANCER RESEARCH

Guest Editor: Julio E. Celis



Neidhardt *et al.* [6] defined the pH gradient in 2-D gel experiments by pI markers whose pI values were calculated from the amino acid composition. Focusing positions of other polypeptides could be predicted from their composition but the pK values needed for the pI calculations were unknown. Various groups employing this approach do not use the same pK values [6, 7] and therefore, the pI values derived in this way cannot be expected to describe the variation of the hydrogen ion activity. In spite of this fact, it is still possible to make approximate predictions of focusing positions because the pK values used to define the pH gradient are also used to calculate pI values and to predict the focusing positions. Errors in pK assignments are therefore compensated. A pH scale which correctly reflects the variation in hydrogen ion activity during focusing should improve the precision of the predictions, but this has never been implemented with CA-IEF focusing as a first dimension in 2-D gel electrophoresis. The main reason for this are the problems associated with pH measurements in focused gels containing high concentrations of urea.

IPGs can be described from the concentration variation of the immobilized groups, provided that the pK values of these groups are known for the conditions prevailing during focusing. To avoid measurements on gels, Gianazza *et al.* [8] suggested the use of pK values derived by addition of determined pK shifts. Recently, direct determinations of pK differences between immobilized groups in IPGs were made by determining pI-pK values in overlapping narrow-range IPGs [9, 10] and the results verified the applicability of the Gianazza approach. A description of the focusing results in a pH scale, which correctly describes the variation of the hydrogen ion activity for the focusing conditions used, not only allows the comparison of 2-D gel maps generated with different IPGs, but also opens the possibility for correlating the focusing position of a polypeptide with its composition [9]. Experiments by Bjellqvist *et al.* [9, 10] have implied that pH scales showing good correlation between calculated and experimental pI values can be derived for any of the conditions commonly used for focusing in connection with 2-D gel electrophoresis. These pH scales are then defined through the pK values of the immobilized groups in the IPG containing gel. To be useful for interlaboratory comparisons, however, the pH scale has to be defined through pI values of easily recognizable spots present in the 2-D gel map. So far, pI determinations in a useful pH scale, combined with determinations of pK values needed for pI calculations, have only been made for the pH range 4.5–6.5 at 10°C [9]. CA-IEF focusing as described by O'Farrell [11] does not control the temperature of the first dimension, which can be expected to be slightly above room temperature. With IPGs, the temperature commonly used is about 20°C [4, 12] or 25°C [13] and this is a critical parameter that needs to be controlled [14].

The present work was designed to compare 2-D gel maps of different cell types in a laboratory applying both CA-IEF and IPG focusing at a common temperature. To this end we have generated 2-D gel maps of proteins from noncultured, unfractionated normal human epidermal keratinocytes with IPG in the first dimension

and a focusing temperature of 25°C. We have used commercial nonlinear, wide-range IPG strips which give 2-D gel maps that are closely similar to the ones resulting with the CA-IEF technique used to establish the human keratinocyte database [15]. As an initial step towards interlaboratory comparisons of results obtained with the nonlinear gradient as a first dimension we report here on the focusing positions of 41 known proteins that are common to most human cell types. The pH range covered corresponds to the range in classical CA-IEF 2-D gel electrophoresis and in order to use these proteins as internal standards for comparing 2-D gel maps generated with other IPGs we determined their pI values with narrow-range IPGs in the first dimension. We have compared the calculated *versus* experimental pI values and show that it is necessary to have further information (absence or presence and nature of posttranslational modifications), in addition to amino acid composition to be able to calculate pI values that correspond to the actual experimental values. The pK values used for the calculations are provided and the usefulness of pI prediction in relation to database information is discussed. Furthermore, we comment on the possibility of using experimentally determined pI values to verify the available database information on polypeptide composition.

2 Materials and methods

2.1 Apparatus and chemicals

Equipment for isoelectric focusing and horizontal SDS electrophoresis (Multiphor® II electrophoresis chamber, Immobiline® strip tray, Multidrive XL programmable power supply, Macrodrive power supply and Multitemp® II) was from Pharmacia LKB Biotechnology AB (Uppsala, Sweden). Vertical second-dimensional gels were run in the home-made equipment described in [15]. The IPG strips with the wide-range nonlinear pH gradient were either Immobiline DryStrip® pH 3–10 NL, 180 mm or alternatively 160 mm long IPG strips with a corresponding pH gradient. In both cases the IPG strips were delivered by Pharmacia LKB. Immobiline, Pharmalyte, Ampholine, GelBond as well as PAG film and the ready-made horizontal SDS gels (ExcelGel® XL SDS 12–14) were also from Pharmacia LKB. Purified proteins and peptides were from Sigma (St. Louis, MO).

2.2 Sample preparation

Preparation and labeling of unfractionated keratinocytes as well as fibroblasts have been described in [16]. Cells were lysed in a solution containing 9.8 M urea, 2% w/v NP-40, 100 mM DTT and 2% v/v Ampholine pH 7–9.

2.3 2-D gel electrophoresis

First-dimensional focusing was performed according to Görg *et al.* [2] with some minor modifications, as described in [9]. Rehydration of the IPG strips was made in a solution containing 9.8 M urea, 2% w/v CHAPS, 10 mM DTT and 2% v/v carrier ampholyte mixture. The carrier ampholyte mixture consisted of 2 parts Pharmalyte

4–6.5. 1 part Ampholine pH 6–8 and 1 part Pharmalyte pH 8–10.5. Usually, cathodic sample application was used and the samples were diluted 2–20 times in a solution containing 9.8 M urea, 4% w/v CHAPS, 1% w/v DTT and 35 mM Tris base. For acidic application, the Tris-base was substituted with 100 mM acetic acid. The degree of dilution and sample volume (20–100 µL) depended on the particular sample and the IPG, and whether visualization of the proteins was to be done by Coomassie Brilliant Blue or silver staining. With the wide-range non-linear IPG, 10–30 µg of total protein was loaded for silver staining and 100–200 µg for Coomassie staining. Focusing was done overnight with Vh products in the range of 45–60 kVh with 160 mm long strips and 50–70 kVh with 180 mm long strips. Solubilization of polypeptides and blocking of -SH groups prior to the second-dimensional run, as well as loading on the second-dimensional gel was done as described in [9]. The stacking gel was omitted and 5–10 mm were left at the top of the second-dimensional gel for applying the IPG strip. The space was filled with electrode buffer containing 0.5% w/v agarose. Casting, running, staining and autoradiography were carried out as described in [15].

2.4 Experimental determination of pI values

The determination of the pK' differences between Immobilines pK' 4.6, pK' 6.2 and pK' 7.0 necessary for the calibration of the pH scale at 25°C in 9.8 M urea was done as described in [9] with the same narrow-range IPGs. The pH scale was defined by setting the pK' value of Immobiline pK' 4.6 equal to 4.61 [9] and the determined pK' differences gave the pK' values of Immobilines pK' 6.2 and pK' 7.0, equal to 5.73 and 6.54, respectively. The pK' differences found are in good agreement with values derived from [17] and [8] by extrapolation to 9.8 M urea concentration. As in [9], additional narrow-range recipes have been used for determining pI values. With narrow-range IPGs extending to pH values higher than the pK' value of Immobiline pK' 7.0, anodic sample application was used with acetic acid added to the sample solution. Otherwise, cathodic sample application was used with the same sample buffer as for wide-range IPGs.

2.5 Protein compositions used for pI calculations

With the exception of vimentin, protein compositions are from the Swiss-Prot database [18]. For vimentin, we used the data from [19], where the amino acid at position 41 is a D instead of a S. Information in the Swiss-Prot database on phosphorylation has been disregarded because it was known from earlier studies (J. E. Celis, unpublished results) that the spots in question corresponded to the unphosphorylated forms of the peptides.

2.6 Calculation of pI values

For the pI calculations it was assumed that the same pK' value could be used for an amino acid residue in all polypeptides and in all positions in the peptide except for N- or C-terminally placed amino acids. For the pK' values of the N-terminal amino groups the effect of the

different substituents on the α-carbon were taken into account. The calculations of pI values were made with the aid of the IPG-maker program [20].

2.7 pK values used for pI calculations

For the carboxyl terminal group and internal glutamyl and aspartyl residues the same pK' values were used as in [9]. For C-terminal glutamyl and aspartyl residues, separate pK' values were derived with the aid of the Taft equations [9, 21]. The pK' values of histidyl groups were calculated from the pI values of human carbonic anhydrase I as in [9]. For N-terminal glycine a pK' value of 7.50 was used. The pK' shift caused by a substituent on the α-carbon was assumed to be identical with the pK' shift the substituent caused for the amino group in the amino acid, i.e. 2.28 pH units were subtracted from the pK' values for the amino groups in the amino acids given in [22, 23]. The approximate pK' value of 9 for the cysteinyl group was taken from [24]. For tyrosyl and arginyl groups we used the pK' values for the amino acids [22, 23]. For lysyl groups the effect of high urea concentration on amino groups was taken into account and 0.5 pH units were subtracted from the amino acid pK' value. These last three pK' values are far from the pH range under study and the results found would have been the same if lysyl and arginyl groups were assumed to be fully ionized while the ionization of tyrosyl groups were neglected. A complete list of the pK' values used is given in Table 1.

Table 1. pK' Values used for the ionizable groups in peptides
9.8 M urea, 25°C

Ionizable group	pK'
C-terminal	3.55
N-terminal	
Ala	7.50
Met	7.00
Ser	6.93
Pro	8.36
Thr	6.82
Val	7.44
Glu	7.70
Internal	
Asp	4.05
Glu	4.45
His	5.98
Cys	9
Tyr	10
Lys	10
Arg	12
C-terminal side chain groups	
Asp	4.55
Glu	4.75

2.8 Statistical analysis

Statistical comparisons of the experimental and calculated pI values were done on an Apple Macintosh IIci using the statistical package Statistica/Mac, release 3.0b (from StatSoft Inc., Tulsa, Oklahoma). Calculated and experimental pI values were compared by the *t*-test for

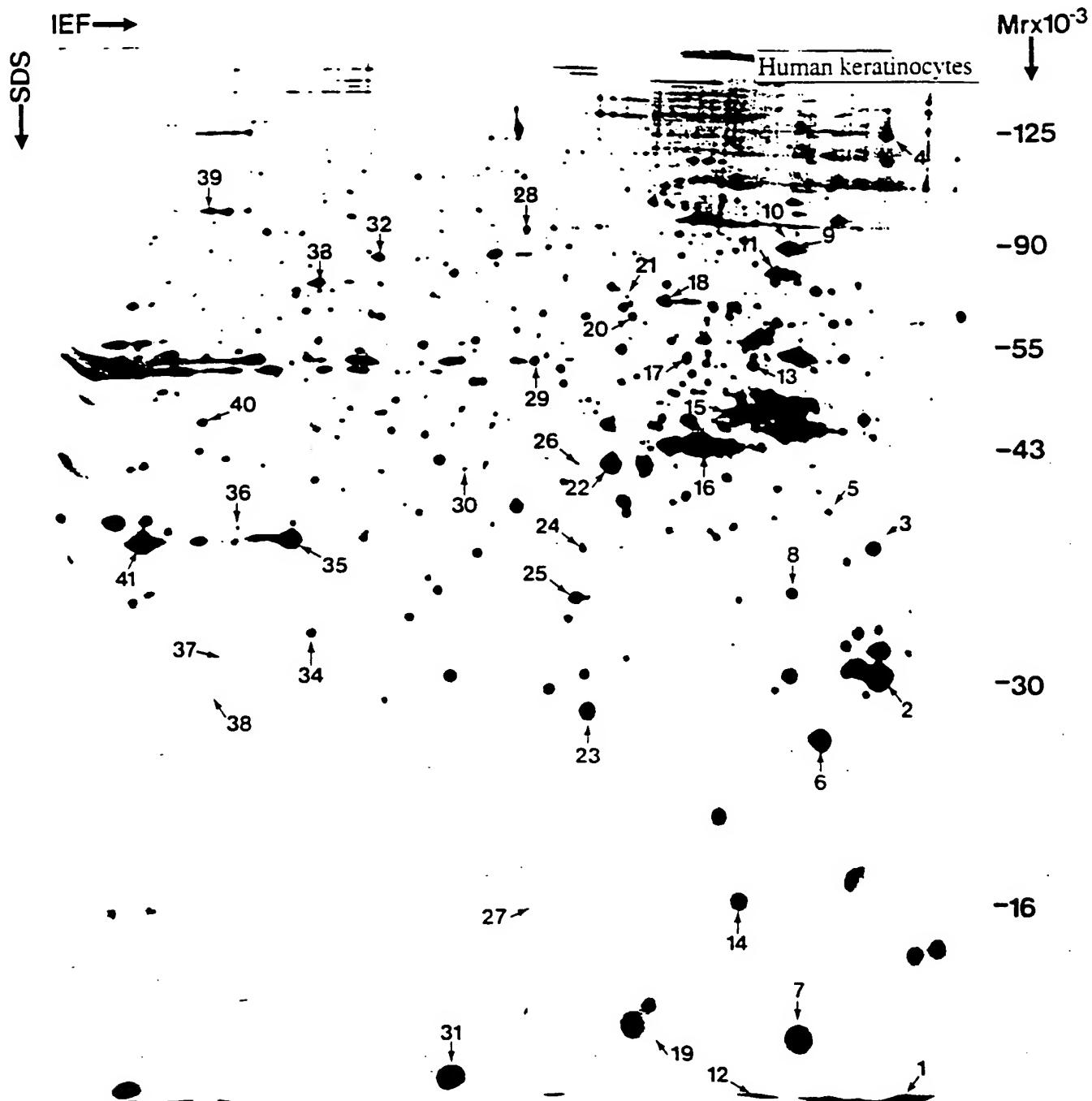
correlated samples (paired *t*-test). The normality of *pI* differences was estimated graphically by probability plots. The variances of the data presented here and the similar data on plasma and liver proteins in [9] were compared by the *F*-test.

3 Results and discussion

3.1 Identification of polypeptides and *pI* determinations

The 2-D gel maps of [³⁵S]methionine-labeled proteins from noncultured, unfractionated normal human kerati-

nocytes, focused with the nonlinear, wide-range IPG and CA-IEF pH gradients in the first dimension, are shown in Figs. 1 and 2, respectively. The IPG extends to higher pH values but otherwise the two patterns are very similar and most of the spots in the IPG pattern can be directly related to the corresponding spots in the CA-IEF gel. To obtain comparable patterns it was important to keep the focusing temperature as similar as possible. Compared to other studies [1–4, 9, 10, 12–14], we increased the urea concentration in the focusing gel to 9.8 M because keratins streaked badly in the focusing dimension when 8 M urea was used, presumably due to



aggregates of acidic and basic keratins. An increase in urea concentration to 9 M or more eliminated these streaks; apart from this effect, no other major changes in the focusing positions were observed. In Fig. 1 we have indicated the positions of 41 known proteins from the human keratinocyte 2-D gel database that are most likely common to most human cell types. The choice was made because these proteins are easy to identify with certainty. With the exception of stratifin (spot 2), involucrin (spot 4) and keratin 14 (spot 15), which are all

epithelial markers, these proteins are also present in human fibroblasts (Fig. 3) and lymphocytes (results not shown), and therefore can be used as landmarks for comparing 2-D gel maps derived from different cell types. In Table 2 the 41 proteins are listed together with their sample spot numbers (SSP) in the human keratinocyte protein database and pI values determined in 2-D gel maps generated with narrow-range IPGs in the first dimension.

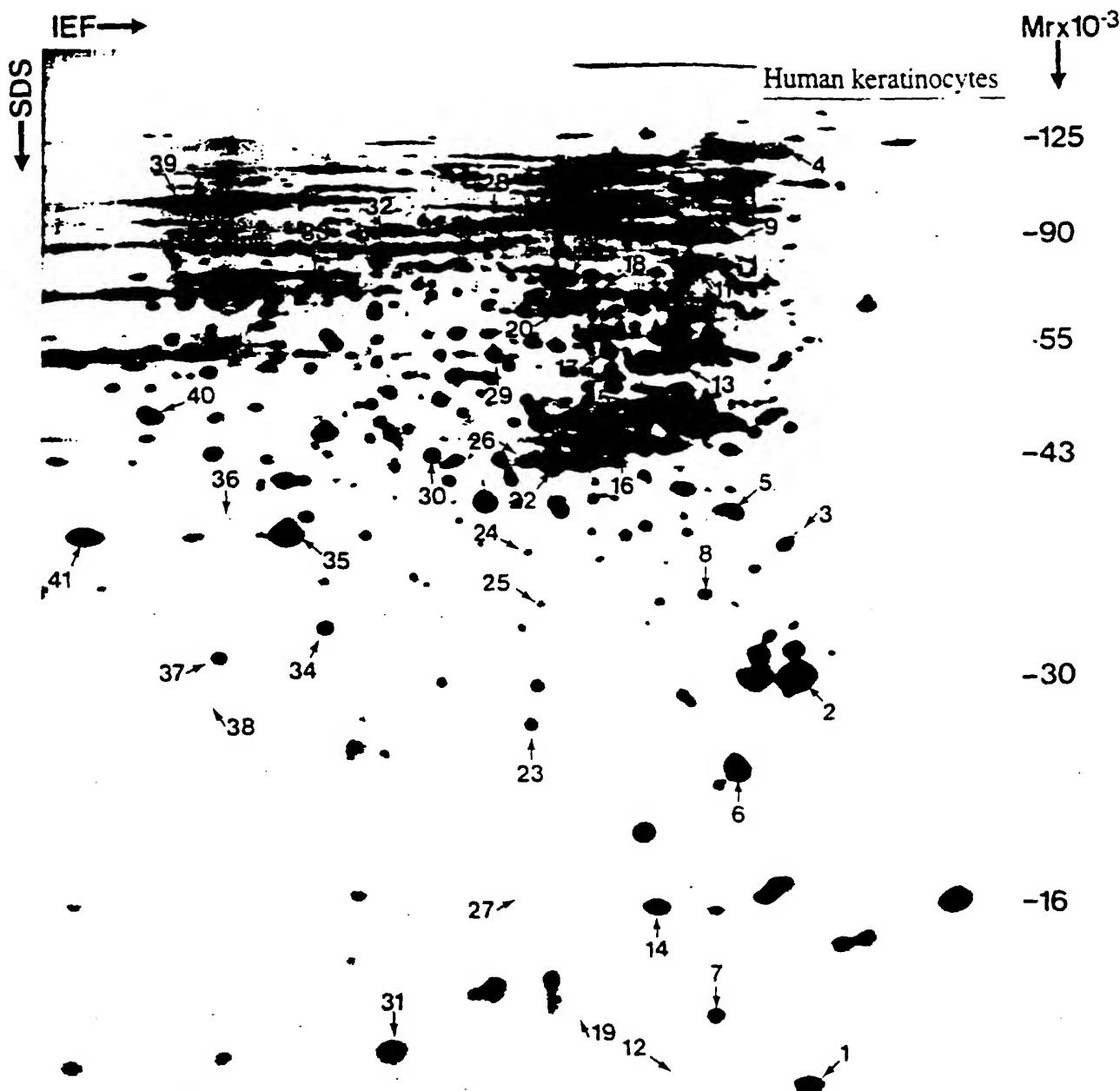


Figure 2. 2-D gel protein map of $[^{35}\text{S}]$ methionine-labeled proteins from noncultured, unfractionated normal human keratinocytes focused with CA-IEF in the first dimension. The position of the 41 proteins analyzed in this study is indicated.

Table 2. Proteins from the human keratinocyte database localized in 2-D gels run with IFCIs as first dimension

Number in Protein name Figs. 1-3	I-1 SSP number ^a	Experimental pI value	Calculated pI value	Discrepancy (pH units)	Buffer capacity		N-terminal blockage accession number
					at experimental pH value	net change at pH value	
1	CatN 19	9027	4.46	-	-	-	
2	Stathmin, bovine 14.3-kDa related protein	9109	4.58	-0.01	-0.1	20.8	N
3	Proliferating nuclear antigen (pCNAP)/cyclin	9226	4.58	-0.01	-0.3	70.1	N
4	Involucrin	9703	4.63	-0.01	-0.3	30.4	N ^b
5	Nucleolar protein B23	8207	4.75	-0.11	-1.2	13.1	N ^b
6	Functionally controlled tumor protein	8114	4.79	-0.05	0.6	7.1	V ^b
7	Thioneoxidase	8006	4.86	-0.04	0.3	20.3	A ^c
8	Annexin V	8213	4.89	-0.01	0.1	56.2	P
9	Heat shock protein 90- β	8611	4.94	-0.01	-0.5	53.6	P
10	Heat shock protein 90- α	2629	4.97	-0.01	0.2	37.5	P
11	Glucose regulated protein 78 (GRP78)	8515	4.99	-0.01	0.6	1	P
12	Cathepsin	8017	5.02	-0.30	-1.3	3.6	M
13	Vimentin	8417	5.05	-0.01	0.2	27.1	S
14	Initiation factor 4D	8016	5.05	-0.03	0.2	7.6	A ^c
15	Keratin 14	7405	5.08	-0.01	0.2	21.0	T
16	β -Actin	7116	5.21	-0.01	0.06	13.3	D ^c
17	Heat shock protein 60	6403	5.23	-0.01	0.1	17.5	A ^c
18	Heat shock cognate 70 kDa protein	6504	5.28	-0.09	1.8	18.1	M
19	Cystatin	6011	5.40	-0.08	0.2	3.0	M
20	1 β -Plastin	6102	5.44	-0.07	1.1	17.7	M
21	Cleactinin	6228	5.45	-0.02	0.5	23.3	N ^b
22	Plasmonogen activator inhibitor-2	6404	5.48	-0.08	0.9	10.7	M
23	Glutathione S-transferase A	5101	5.43	-0.01	0.08	3.9	P
24	Annexin VIII	5213	5.45	-0.11	1.0	8.7	M
25	Annexin III	5204	5.46	-0.17	1.4	8.4	M
26	Adenosine deaminase	5105	5.47	-0.16	1.8	10.8	M
27	Stathmin	5001	5.55	-0.61	0.4	6.6	N ^b
28	Gelsolin, cytoplasmic	5608	5.59	-0.58	0.1	16.5	V
29	Rat phosphomoside specific protein homolog	5110	5.62	-	-	-	
30	Filastase inhibitor	4314	5.74	-	-	-	
31	Slc100, calgratin	4006	5.75	-	-	-	
32	Cavulin, cavin	3501	5.99	-0.01	0.5	13.2	P
33	Miasin	1515	6.11	-0.09	0.2	9.8	
34	Purine nucleoside phosphorylase	2108	6.11	-0.15	1.8	4.1	M
35	Annexin I	2246	6.18	-0.64	0.46	2.5	A
36	Adrose reductase	1202	6.40	0.55	0.15	0.7	A
37	Phosphoglycerate mutase (B form)	1007	6.46	0.75	0.29	0.9	A ^c
38	Triosephosphate isomerase	1111	6.53	0.51	-0.02	2.3	N
39	Elongation factor 2	1610	6.43	0.38	0.05	0.5	N
40	α -Endonuclease	1325	6.62	0.99	0.37	1.0	S
41	Annexin II	210	7.30	7.36	0.06	0.05	0.9

^a SSP number in the keratinocyte database [15]

^b Peptides N-terminally sequenced as liver proteins [3]

^c Peptides given as N-terminally blocked in Swiss-Prote database

3.2 Comparison between the determined and calculated pI values for human keratinocyte proteins

Thirty six of the 41 proteins listed in Table 2 are found in the Swiss-Prot database. Contrary to the plasma and liver proteins used in [9], the pI calculations on the proteins used in this study posed some problems that reflected the way in which they were characterized. The

proteins used by Bjellqvist *et al.* [9] were either very abundant and well-characterized plasma proteins or they were identified by N-terminal sequencing and, therefore, the nature of the N-terminals (acetylated or non-acetylated) was in both cases known. The proteins used in this study have all been characterized by internal sequencing [7] and it is known that N-terminal acetylation occurs with high frequency in eukaryotes.

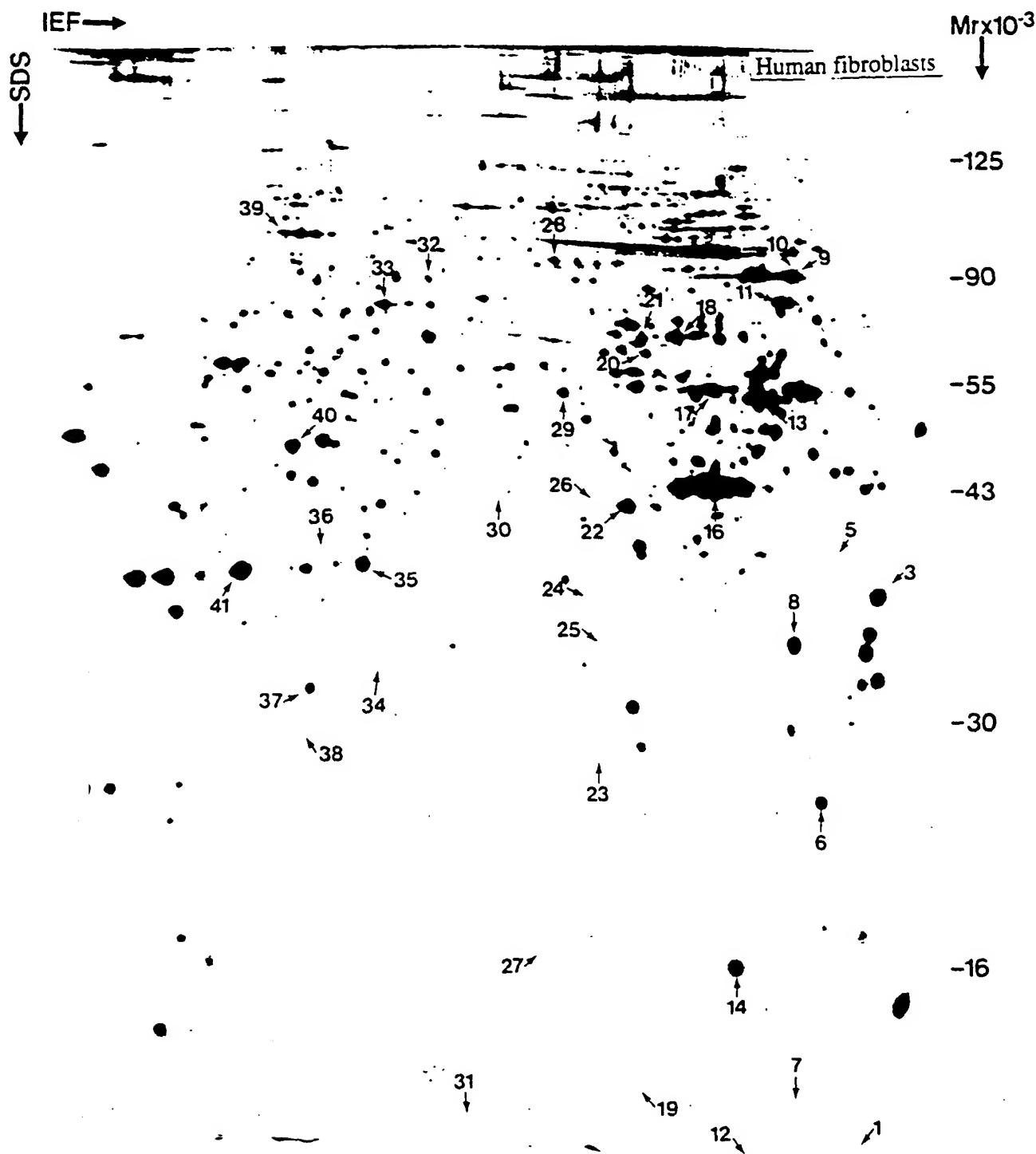


Figure 3. 2-D protein map of $[^{35}\text{S}]$ methionine-labeled proteins from normal human fibroblasts focused with the nonlinear, wide-range IPG in the first dimension. The position of the 41 proteins analyzed in this study is indicated.

According to Brown and Robert [25], proteins with acetylated *N*-terminals correspond in weight to approximately 80% of the soluble protein in ascites cells. Based on results from *N*-terminal sequencing, at least 40% of the spots in the human liver protein 2-D gel map appear to be blocked [3]. The corresponding number, derived from 107 spots in the 2-D gel map of human T-lymphocyte proteins, falls between 60 and 65% (J. Strahler, personal communication). Information concerning *N*-terminal blockage is not normally available, and in the Swiss-Prot database only 6 of the 36 keratinocyte proteins are specified as *N*-terminally blocked. We have, within the present material, defined 18 proteins for which the *N*-terminals are very likely to be correctly described. Six of these proteins are listed in the Swiss-Prot database as *N*-terminally blocked, four represent proteins which appear in the human liver 2-D gel map and have been *N*-terminally sequenced as liver proteins [3] and the remaining eight have *N*-terminal groups other than M, S and A, *i.e.* *N*-terminals for which *N*-acetylation is uncommon [26]. In Figs. 4A, B, C and D *p*/*I* values calculated from Swiss Prot database information are plotted against the experi-

mentally determined *p*/*I* values for all the keratinocyte proteins listed in Table 2 and for the 18 selected proteins, as well as for the plasma and liver proteins (data from [9] valid for 10°C)*.

The calculations show that without knowledge of the status of the *N*-terminal group, precise predictions of *p*/*I* values for eukaryotic proteins cannot be achieved based on the information available in Swiss-Prot and similar databases. However, for proteins where the *N*-terminal status is known, we find good correlation between predicted and experimental *p*/*I* values. When the variance of the *p*/*I* discrepancies and the variance of calculated charges at the experimental *p*/*I* values derived from the present data set are compared with the corresponding

* There are four plots: (A) the 36 polypeptides from normal human keratinocytes (no corrections), (B) the 36 polypeptides from Fig. 4A where *p*/*I* values have been recalculated for 12 polypeptides with M, S and A as *N*-terminally assumed blocked, based on calculated charge, (C) the 18 selected polypeptides with information on the *N*-terminal configuration, and (D) plasma and liver proteins.

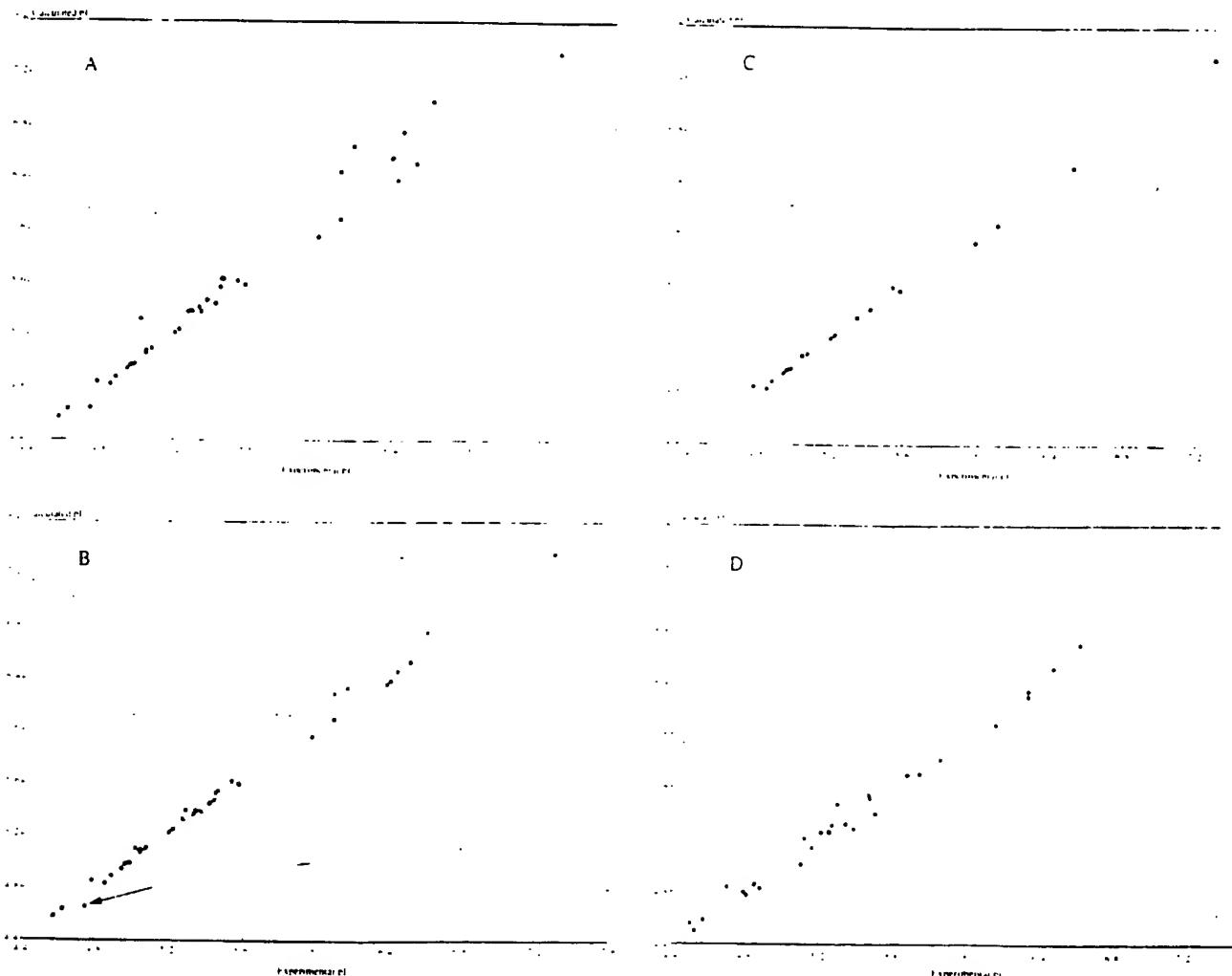


Figure 4. Calculated vs. experimental *p*/*I* values. Lines are fitted using the least squares' criterion. (A) 36 polypeptides from normal human keratinocytes (no corrections), (B) 36 polypeptides from Fig. 4A (including the 18 marker polypeptides) where *p*/*I* values have been recalculated assuming *N*-terminal blockage; x indicates recalculated *p*/*I* values; nucleolar protein B23 is indicated with an arrow. (C) 18 polypeptides with information on *N*-terminal configuration and (D) plasma and liver proteins.

values derived from the data on plasma and liver proteins in [9] (Table 3), the present data are found to result in larger variances for the values of both pI discrepancies and calculated charge at the experimental pI value when no information on posttranslational modification is taken into consideration. Correction for possible N-acetylation of 12 polypeptides with M, S and A as N-terminal results in a smaller variance of pI discrepancies, although not significantly different from values derived from [9], whereas the variance of the calculated charge at the experimental pI value is significantly higher. For the 18 selected proteins the variance for the pI discrepancies is significantly smaller than for the data in [9]; however, the corresponding value for calculated charge at the experimental pI value does not improve to the same extent. This, we believe, reflects another difference between the two sets of proteins used for the calculations. Based on spot distributions in 2-D gel maps, the set of proteins used here has a molecular weight distribution that is more representative of the patterns observed in mammalian cells. In the study by Bjellqvist *et al.* [9] most of the high molecular weight plasma proteins had to be excluded due to their unknown content of sialic acid which made the proteins analyzed in this study heavily biased towards low molecular weight proteins. The buffer capacity of proteins normally increases with the protein's molecular weight, and the average buffer capacity of the presently selected proteins with assumed known N-terminals is 18 charge units/pH unit, while the corresponding value for the proteins used in [9] is only 9 charge units/pH unit. High buffer capacity can be expected to improve the agreement between calculated and experimental pI values. Inspection of the data presented in Table 2 for the polypeptides with assumed known N-terminals verifies the importance of the buffer capacity. For 8 polypeptides having buffer capacities higher than 15 charge units/pH unit, the calculations in all cases yielded pI discrepancies with absolute values of less than 0.02 pH units. The largest discrepancy, 0.06 pH units, was observed for annexin II and stathmin, proteins which have low buffer capacity: 0.9

and 6.6 charge units/pH unit, respectively. The probability that the focusing position of a protein with known composition will fall within a certain distance from the calculated pI value therefore cannot be predicted by the variance alone. The buffer capacity of the specific protein must be taken into consideration as well. As indicated by the decrease of the variance of calculated charges at the experimental pI value for the selected proteins, the observed improvement can not solely be due to the higher buffer capacity of the keratinocyte proteins. The two studies relate to different experimental conditions. Good agreement between experimental and calculated pI values implies that the proteins are defolded and a factor that may contribute to the observed improvement is a more complete defolding of proteins caused by the higher temperature and urea concentration used in this study.

The data indicated that the precision with which pI values can be predicted for polypeptides with high buffer capacity is better than the precision with which experimental pI values can be determined. If the pH is defined through the pK values of the immobilized groups in the IPG containing gel, the precision of the experimentally calculated data will depend on the pH difference between the pI and the pK value of the immobilized group with the closest pK. For the present study this will give pI determinations with a precision varying in the range of ± 0.02 – 0.05 pH units [9]. The good agreement observed between the calculated and experimental pI values is due to the fact that errors are mainly systematic and, as discussed in [9], they will largely be cancelled out in the calculations. A pH scale defined through the presently determined pI values will not necessarily reflect the variation of the hydrogen ion activity during the focusing step in an optimal way, but it still allows precise predictions of focusing positions for polypeptides with known compositions, including information on posttranslational modifications. Calculated net charge at the experimentally found isoelectric point defined in this scale will serve as a tool to verify that the polypeptide

Table 3. Mean values and variances for the difference (experimental pI-calculated pI) in pH units and calculated charges at the experimental pI values, respectively

Number of proteins	Plasma and liver proteins (18 M urea, 10°C)				Keratinocyte proteins (9.8 M urea, 25°C)			
	All peptides		All peptides after correction for N-acetylation		Known N-terminal configuration (or very likely configuration)		18	
	29	36	36					
Experimental pI-calculated pI	Mean -0.011	Variance 0.005	Mean 0.072	Variance 0.017	Mean 0.019	Variance 0.003	Mean 0.005	Variance 0.001
F-value (pI discrepancy) ^a	1		3.4		1.67		5	
P-level (pI discrepancy) ^a	0.5		0.0005		0.0721		0.0004	
Calculated charge at the experimental pI value	-0.070	0.22 ^b	0.321	0.871	0.009	0.444	-0.014	0.109
F-value (calculated charge at the experimental pI value) ^a	1		3.8		1.96		2.08	
P-level (calculated charge at the experimental pI value) ^a	0.5		0.0002		0.0338		0.0536	

^a Comparison to the data in [9]. $F = S_1^2/S_2^2$, where S_i^2 is the larger of the two variances

^b $P(F(v_1, v_2) \geq F\text{-value})$, where v_1 and v_2 are the degrees of freedom for s_1 and s_2 , respectively

composition used in the calculation is correct and complete. Exceptions to this are proteins such as involucrin and heat shock protein 90 that have very high buffer capacities. Introduction of an extra charge unit into these proteins will only result in *pI* shifts falling in the range of 0.01–0.02 pH units and the effect is that the quality of the pH definition – the precision by which *pK* values used in the calculations are given and the precision of experimental *pI* values in these cases – will limit the possibilities to verify polypeptide composition based on the experimental *pI* value.

Statistical comparison of experimental and calculated *pI* values was done using the *t*-test for dependent samples and normality of the discrepancies was estimated by probability plots. For the 36 proteins, the *p*-level is 0.0021, indicating that a result like this is unlikely to be a chance effect and must be assumed to represent a real difference. After correction for the most likely *N*-terminal configuration, the *p*-level is 0.043 and cannot be accepted as representing the same population since the *p*-level is less than 0.05 – the traditional *p*-limit of statistical significance. For the 18 proteins with a known or very likely *N*-terminal configuration the *t*-test gave a *p*-level of 0.49, which verifies that the experimental and calculated *pI* values are not significantly different.

Besides showing that *pI* values for denatured proteins with known compositions can be calculated with a high degree of precision from average *pK* values, the results also provide strong support for the notion that *N*-terminal blockage heavily depends on the nature of the *N*-terminal groups [26]. The results seem to indicate that with *N*-terminals other than M, S and A, only a few proteins have blocked *N*-terminals (1 out of 10 proteins in the present study), while it can be inferred from the data presented in Table 2 that a majority of the proteins with M, S and A as *N*-terminal are blocked. After correction for the effect of suspected *N*-terminal blockage there is only one protein (nucleolar protein B23) out of the 36 used in this study, which, in spite of a high buffer capacity, has a marked difference of 0.11 pH units between predicted and determined *pI* values (Fig. 4B); this corresponds to 3 charge units due to the high buffer capacity of this protein. This discrepancy in *pI* prediction and calculation of net charge at the *pI* is probably not due to deficiencies in the database information but instead reflects a shortcoming of the model used for *pI* calculations. Nucleolar protein B23 contains a domain extremely rich in aspartic and glutamic acid residues (Table 4), in which 26 out of 28 amino acid residues from position 161 to 188 are either a D or an E. A calculation based on the use of average *pK* values uninfluenced by the charged neighboring amino acid residues cannot be expected to correctly describe the *pI* value with almost half of the acidic groups packed

together into a highly negatively charged region. This limitation caused by calculations based on average *pK* values does not severely limit the usefulness of the approach since a search through Swiss-Prot shows that this type of D/E-rich motif is uncommon, and the existence of a highly charged region is immediately apparent upon inspection of the amino acid sequence.

The quality of the information available in databases, especially concerning posttranslational modifications, is a major problem when the data is to be used for *pI* predictions. The *p*-level of 0.043 found for all 36 proteins after correction for *N*-acetylation, shows that this problem is not only limited to *N*-terminal blockage and the very good agreement found for the eighteen polypeptides, with assumingly correctly described *N*-terminal (Fig. 4C), must be regarded as an exception from this point of view. *N*-Terminal blockage is generally the main problem in relation to *pI* predictions for eukaryotic proteins. Of the 36 keratinocyte proteins analyzed, 18–20 are suspected to be *N*-terminally blocked (6 proteins blocked according to Swiss-Prot, 12 proteins with M, S or A as *N*-terminal and assumingly blocked based on the calculated charge, and two proteins, involucrin and nucleolar protein B23, with M as *N*-terminal for which the data does not allow any conclusion). This is in reasonable agreement with the conclusions based on the *N*-terminal sequencing data derived in connection with 2-D gel electrophoresis. *N*-terminal blockage can be suspected for 17–19 of the 26 proteins with M, S or A as *N*-terminal, while only 1 in 10 proteins with other *N*-terminal groups are blocked. The information that the frequency of *N*-terminal blockage is strongly related to the nature of the *N*-terminal group will be of some help in connection with *pI* predictions based on database information. However, without information from other sources, an uncertainty will always remain as to whether the *N*-terminal charge should be included in the *pI* calculation.

4 Concluding remarks

The data presented here lays the foundation for comparing 2-D gel protein maps of different cell types generated with nonlinear, wide-range IPGs in the first dimension. The focusing positions of 41 polypeptides common to most human cell types have been described in a pH scale that allows focusing positions to be predicted with a high degree of accuracy, provided that the composition of the polypeptides are known and that information on posttranslational modifications are available. For polypeptides with a very high buffer capacity, the limiting factor is the precision with which experimental pH values can be determined rather than the precision of the calculations. Possible deficiencies in the pH scale description of the variation of the hydrogen ion activity has, at least at the present state, no consequences for its practical use. The major limitation in connection with predictions of focusing positions from polypeptide compositions is the quality of existing data on protein compositions, especially concerning posttranslational modifications. Amino acid sequences have been reasonably easy to obtain, while posttranslational modifications

Table 4. Amino acid sequence of nucleolar phosphoprotein B23

1	MEDSQQMDS	PFLPQPLPS	CEPQDQDQH	PAEAEDEEH	PLSPLTFLSD
51	AGAQNDEEHTY	EAEPEESESS	PDQPLATLH	WVTPWHDG	SPECIFPFAL
101	FLLQGSSSFTH	ASQQLAATE	EAEPEESESS	ETPPLSDDG	KPAAPODPTY
151	VPPKRTYLAA	EEPEESESS	EEPEESESS	EEPEESESS	PTPSSRDPDF
201	AATRCKEYQH	WDEKQFSESTP	RKQDQEFTK	QETPPTWHS	PSSTEEHDK
251	WQASIEENGG	LDTPEEFTD	WVTPWHDG	TEPPTQDQW	PKEL

have been difficult and work-intensive to determine. Recent developments in the field of mass spectrometry are fast changing this situation and within the next years we can expect a surge in reliable data in this area. While awaiting this development, verification of correctness and completeness of available information on polypeptide composition can be provided by experimental pI values in a pH scale based on the pI values determined in this study. So far, our data cover the pH range below pH = 7.5. The basic pH range covered by NEPHGE as first dimension will be covered in forthcoming work.

Received December 29, 1993

5 References

- [1] Gianazza, E., Astrua-Testori, S., Caccia, P., Giaccon, P., Quaglia, L., Righetti, P. G., *Electrophoresis* 1986, **7**, 76-83.
- [2] Görg, A., Postel, W., Günther, S., *Electrophoresis* 1988, **9**, 531-540.
- [3] Hochstrasser, D. F., Frutiger, S., Paquet, N., Barroch, A., Ravier, F., Pasquali, C., Sanchez, J.-C., Tissot, J.-D., Bjellqvist, B., Vargas, R., Appel, R. D., Hughes, G. J., *Electrophoresis* 1992, **13**, 992-1001.
- [4] *Immobiline DryStrip Kit for 2-D Electrophoresis: Instructions*. Pharmacia LKB Biotechnology AB, Uppsala 1993.
- [5] Anderson, N. L., Hickman, B. J., *Anal. Biochem.* 1979, **93**, 312-320.
- [6] Neidhardt, F. C., Appleyby, D. A., Sankar, P., Hutton, M. E., Phillips, T. A., *Electrophoresis* 1989, **10**, 116-121.
- [7] Rasmussen, H. H., Dumme, J. V., Puypé, M., Gesser, B., Celis, J. E., Vandekerckhove, J., *Electrophoresis* 1992, **13**, 960-969.
- [8] Gianazza, E., Artoni, G., Righetti, P. G., *Electrophoresis* 1983, **4**, 321-326.
- [9] Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.-C., Frutiger, S., Hochstrasser, D. F., *Electrophoresis* 1993, **14**, 1023-1031.
- [10] Bjellqvist, B., Pasquali, C., Ravier, C., Sanchez, J.-C., Hochstrasser, D. F., *Electrophoresis* 1993, **14**, 1357-1365.
- [11] O'Farrell, P. H., *J. Biol. Chem.* 1975, **250**, 4007-4021.
- [12] Görg, A., *Biochem. Soc. Transactions* 1993, **21**, 130-132.
- [13] Hanash, S. M., Strahler, J. R., Neel, J. V., Haitat, N., Malhem, R., Keim, D., Zhu, X. X., Wagner, D., Gage, D. A., Watson, J. T., *Natl. Acad. Sci. USA* 1991, **88**, 5709-5713.
- [14] Görg, A., Postel, W., Friedrich, C., Kuick, R., Strahler, J. R., Hanash, S. M., *Electrophoresis* 1991, **12**, e53-e58.
- [15] Celis, J. E., Rasmussen, H. H., Olsen, E., Madsen, P., Leffers, H., Honore, B., Dejgaard, K., Gromov, P., Hoffmann, H. J., Nielsen, M., Vassilev, A., Vintermyr, O., Hao, J., Celis, A., Basse, B., Lauridsen, J. B., Ratz, G. P., Andersen, A. H., Walbum, E., Kiærgaard, L., Puypé, M., Van Damme, J., Delay, B., Vandekerckhove, J., *Electrophoresis* 1993, **14**, 1091-1198.
- [16] Celis, J. E., Madsen, P., Rasmussen, H. H., Leffers, H., Honore, B., Gesser, B., Dejgaard, K., Olsen, E., Magnusson, N., Kil, J., Celis, A., Lauridsen, J. B., Basse, B., Ratz, G. P., Andersen, A., Walbum, E., Brandstrup, B., Pedersen, P. S., Brandt, N. J., Puypé, M., Van Damme, J., Vandekerckhove, J., *Electrophoresis* 1991, **12**, 802-812.
- [17] Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Görg, A., Postel, W., Westermeier, R., *J. Biochem. Biophys. Methods* 1982, **6**, 317-333.
- [18] Barroch, A., Boeckman, B., *Nucleic Acids Res.* 1991, **19**, 2247-2250.
- [19] Honore, B., Madsen, P., Basse, B., Andersen, A., Walbum, E., Celis, J. E., Leffers, H., *Nucleic Acids Res.* 1990, **18**, 6692.
- [20] Altland, K., *Electrophoresis* 1990, **11**, 140-147.
- [21] Perrin, D. D., Dempsey, B., Serjeant, E. P., *pKa Predictions for Organic Acids and Bases*. Chapman and Hall Ltd., London 1981.
- [22] Perrin, D. D., *Dissociation Constants of Organic Bases in Aqueous Solutions*. Butterworths, London 1965.
- [23] Perrin, D. D., *Dissociation Constants of Organic Bases in Aqueous Solutions*. Supplement 1972. Butterworths, London 1972.
- [24] Altland, K., Becher, P., Rossman, U., Bjellqvist, B., *Electrophoresis* 1988, **9**, 474-485.
- [25] Brown, J. L., Robert, W. K., *J. Biol. Chem.* 1976, **251**, 1009-1014.
- [26] Persson, B., Flinta, C., Heine, G., Jornvall, H., *Eur. J. Biochem.* 1985, **152**, 523-527.

Proprietary Info

LSB & LSP Information

Large Scale Biology Corporation

Large Scale Proteomics Corporation

Large Scale Biology Corporation

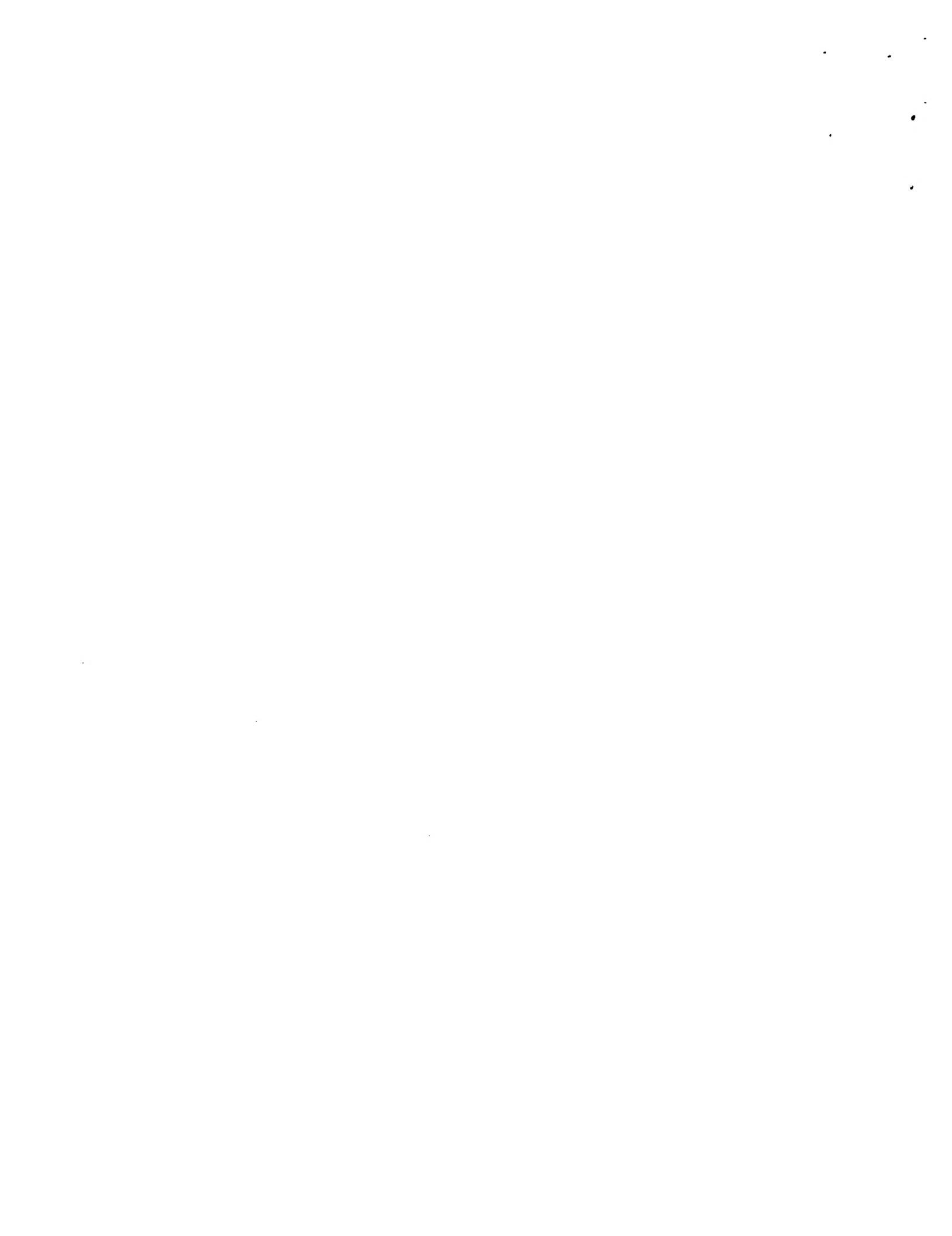
Large Scale Biology Corporation is the leader in the integrated discovery, production and application of proteins - the functional units of all biological processes.

Large Scale Biology Corporation (LSB, Vacaville, CA) and its subsidiary Large Scale Proteomics Corp. (LSP, Germantown, MD) are a biotechnology enterprise with the mission of accelerating the speed and productivity of the life sciences industry product discovery and development programs. Unique among biotechnology companies is LSB's integration of technologies to discover, analyze, manufacture and find new applications for proteins - the functional units of all biological processes.

Genomics companies have focused on deciphering genetic information, providing an initial but only partial understanding of biological processes. LSB's proprietary protein technologies can enable the transformation of genomic information into products such as drug targets, therapeutics, diagnostics for drug efficacy and toxicity, and traits for agricultural crops. Large Scale Biology has gone beyond the "genomics" realm in its business model and developed ways to integrate the discovery of gene function with quantitative protein analysis and protein manufacturing. This integration of technology platforms favorably positions LSB as a leading provider of valuable content to industry leaders in the fields of diagnostics, therapeutics, vaccines and agribusiness.

LSB was founded in 1987 with the goal of commercializing its proprietary GENEWARE viral vector system - a novel technology for gene expression. Using safe RNA viruses to transiently express genes in non-recombinant plants, LSB has positioned itself in the industry to provide cost-effective manufacturing and purification of diverse protein and peptide products. The same technology can be applied to the expression of libraries of foreign genes in an automated, high-throughput format to discover the function of genes with unparalleled efficiency. The GENEWARE system and associated proprietary technologies form the basis for LSB's functional genomics, biomanufacturing and a variety of proprietary products under development.

From its foundation, LSB understood the need to integrate functional genomic and protein manufacturing expertise with quantitative protein analysis and informatics to become a world-leader in the protein field. In 1999, LSB acquired a privately held pharmaceutical proteomics company originally founded in 1985. Large Scale Proteomics Corporation (a wholly



owned subsidiary of Large Scale Biology Corporation) is an industry leader in identifying and characterizing proteins in all types of biological samples for the discovery and development of new and more effective therapies, diagnostics, and agricultural products.

"Proteomics" is the study of the entire complement of proteins expressed in a cell, tissue, or organism. Proteomics can significantly improve drug discovery and development because most illness is associated with imbalances among, or malfunctions of, proteins. Only a small fraction of diseases can be attributed to the presence of a defective gene. Unlike classical genomics approaches that discover genes that may relate to a disease, LSP has developed a proprietary system called the ProGEx module for directly characterizing proteins associated with disease. Using this same technology, LSP can characterize the effects of candidate drugs intended to reverse a disease process, and to determine the degree to which this objective is achieved free of adverse side effects.

LSB and LSP have protected their many discoveries through an extensive portfolio of domestic and foreign patents and have developed commercial alliances and partnerships to exploit the value of their technologies. LSB and LSP scientists and engineers focus on the development and application of resources to help clients meet their objectives as well as the development of our own proprietary products for subsequent partnering with industry leaders.

A combined staff of 140 professionals operates from three locations in the United States, with a network of collaborators and affiliates throughout the US and Europe. Company headquarters, R&D laboratories and its Genomics division are located in Vacaville, California about 60 miles northeast of San Francisco. Process development and biomanufacturing take place in Owensboro, Kentucky, and LSB's Large Scale Proteomics Corporation subsidiary is located in Germantown, Maryland.

In August, 2000, LSB completed an initial public offering (IPO) of 5 million shares of common stock and now trades on the NASDAQ under the symbol LSBC.

Leadership - Large Scale Biology Corporation

Robert L. Erwin, Chairman of the Board and Chief Executive Officer, founded LSB™ and has served as a director and officer since 1987. Mr. Erwin is the former chairman of the State of California Breast Cancer Research Council and currently serves on the University of California President's Engineering Advisory Council. He is Chairman of the Supervisory Board of Icon Genetics AG. As a co-founder of Sungene Technologies Corp., Mr. Erwin served as Vice President of Research and Product Development from 1981 through 1986. He has served on the Biotechnology Industry Advisory Board for Iowa State University. Mr. Erwin received his M.S. degree in Genetics from Louisiana State University and is an inventor on several LSB patents.

David R. McGee, Ph.D., a co-founder of LSB and Senior Vice President and Chief Operating Officer, has been an officer since 1987. Prior to joining LSB, Dr. McGee was Vice President of Operations at Sungene Technologies Corporation from 1983 to 1987. Dr. McGee received his Ph.D. in Genetics from Louisiana State University and served as a faculty instructor of zoology and genetics at Louisiana State University.

Laurence K. Grill, Ph.D., a co-founder of LSB and Senior Vice President, Research and Development, has served as an officer since 1987. Dr. Grill was the Manager of Plant Molecular Biology for Sandoz Crop Protection Corp. from 1984 to 1987 and Senior Research

Scientist in the Department of Molecular Biology at Zoecon Research Institute from 1980 to 1984. He received his Ph.D. from the University of California at Riverside with an emphasis on the molecular basis for viral gene expression in plants.

R. Barry Holtz, Ph. D., Senior Vice President, Biopharmaceutical Manufacturing, has served the company as an officer since 1989 upon the acquisition of Holtz Bio-Engineering, which was founded in 1980. Dr. Holtz was a co-founder and Director of Research for MFI, Inc., the largest manufacturer of microencapsulated nutrients for agriculture and Director of Fundamental Research at Foremost-McKesson, Inc. Dr. Holtz received his Ph.D. in Biochemistry from Pennsylvania State University and served as Assistant Professor in the Department of Food Science and Nutrition at Ohio State University.

Daniel Tusé, Ph.D., has been an officer of LSB since he joined the Company in 1995 as Vice President, Pharmaceutical Development. Dr. Tusé manages the company's pharmaceutical design and development programs, including LSB's novel vaccines and immunotherapeutics initiatives. Prior to joining LSB, Dr. Tusé was Assistant Director of SRI International's (Menlo Park, Calif.) Life Sciences Division. In his 17 years at SRI, Dr. Tusé developed extensive R&D experience in pharmaceuticals and specialty chemicals, serving an international list of clients. Dr. Tusé received his Ph.D. in Microbiology (1980, *cum laude*) with a minor in Toxicology from the University of California, Davis.

John S. Rakitan, a co-founder of LSB, Senior Vice President & General Counsel and Secretary, has served as an officer since 1988. Prior to joining LSB, Mr. Rakitan was an attorney in private practice. Mr. Rakitan received his J.D. degree from the University of Notre Dame.

Michael D. Centron, Treasurer, has served as Controller since 1988 and was elected as Treasurer in 1991. Mr. Centron was Audit Supervisor for Varian Associates from June 1985 through July 1988, and he also worked for Arthur Young and Co. (currently Ernst & Young). Mr. Centron is a certified public accountant and received his M.B.A. degree from the University of California at Berkeley.

Guy della-Cioppa, Ph.D., is an officer of the company and currently serves as Vice President, Genomics. Prior to joining the company in 1989, Dr. della-Cioppa worked for Monsanto Company in St. Louis, MO from 1984-1989 and was an NIH Postdoctoral Fellow at the Worcester Foundation for Experimental Biology in Shrewsbury, MA from 1983-1984. He received his Ph.D. in Biology from the University of California, Los Angeles.

William M. Pfann joined Large Scale Biology in August 2000 as Senior Vice President Finance and Chief Financial Officer. Mr. Pfann was formerly with PricewaterhouseCoopers LLP from 1969 to July 2000, most recently as the Risk Management Partner for the Western Region. He served in a number of management roles at PwC, including leader of the firm's Silicon Valley audit practice, National Director of the networking and communications sector and Managing Partner of the Northern California emerging business group, as well as Partner-in-Charge of the Oakland and Walnut Creek, California offices. Mr. Pfann received a B.S. degree from the University of California, Berkeley, in Business Administration and an MBA in Accounting from Golden Gate University.

[back to index](#)

Large Scale Proteomics Corporation

Leadership - Large Scale Proteomics Corporation

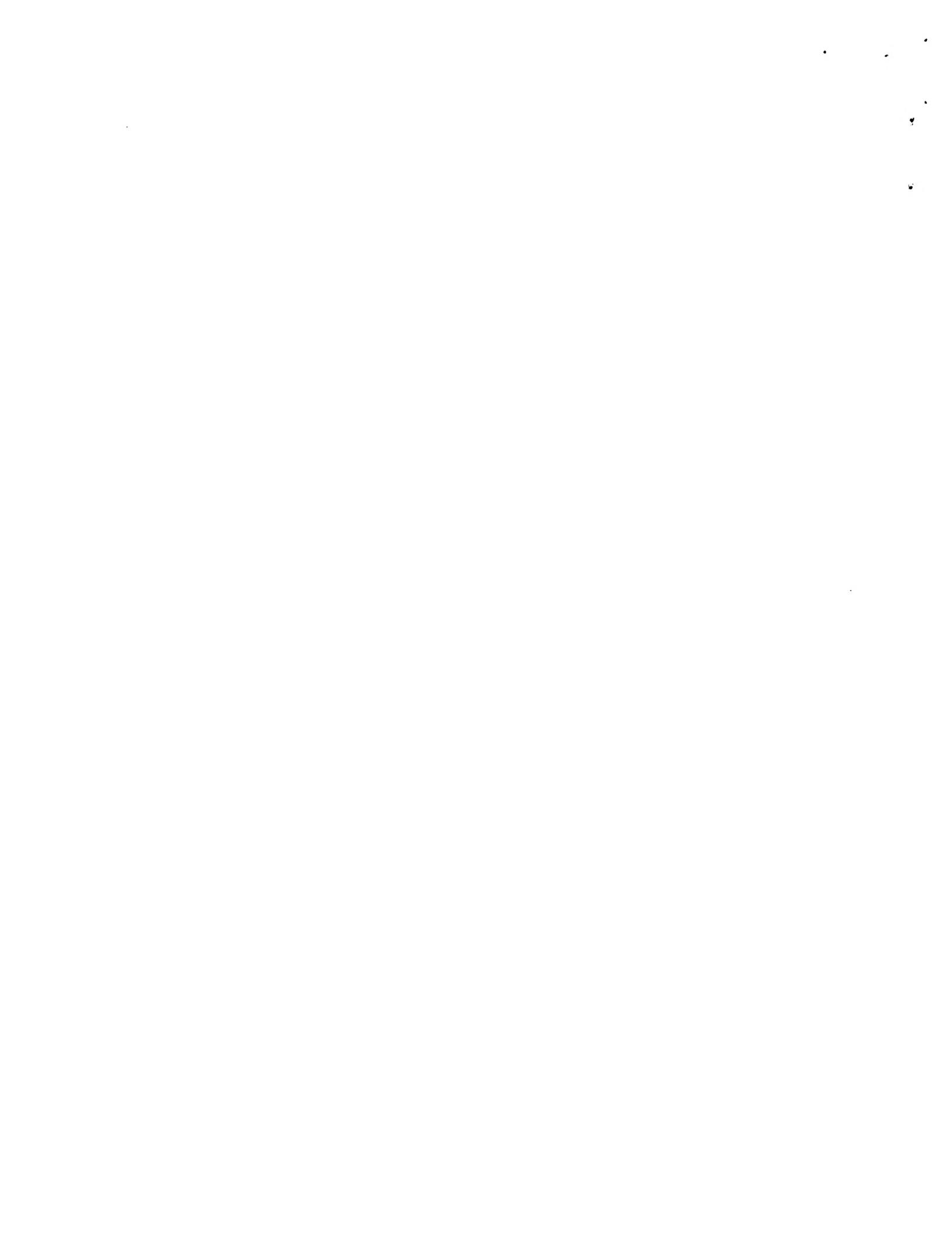
N. Leigh Anderson, Ph.D., Chairman, President and CEO of Large Scale Proteomics Corporation (LSP™). Dr. Anderson obtained his B.A. in Physics with honors from Yale and a Ph.D. in Molecular Biology from Cambridge University (England) working with M. F. Perutz as a Churchill Fellow at the MRC Laboratory of Molecular Biology. Subsequently he co-founded the Molecular Anatomy Program at the Argonne National Laboratory (Chicago) where his work in the development of 2-dimensional electrophoresis (2-DE) and molecular database technology earned him, among other distinctions, the American Association for Clinical Chemistry's Young Investigator Award for 1982 and the 1983 Pittsburgh Analytical Chemistry Award. In 1985 Dr. Anderson co-founded LSP (originally Large Scale Biology Corp., Germantown, MD) in order to pursue commercial development and large-scale applications of 2-D electrophoretic protein mapping technology.

Norman G. Anderson, Ph.D., Chief Scientist at LSP. Dr. Anderson has a distinguished record as an inventor. His career includes senior positions at Oak Ridge and Argonne National Laboratories (ORNL and ANL), more than 300 scientific publications, and the receipt of more than 20 prestigious awards in recognition of his work in science and technology. For his invention of the zonal ultracentrifuge, he received the John Scott Medal Award, and for the centrifugal fast analyzer, the Preis Biochemische Analytik für Klinische Chemie from Die Deutsche Gesellschaft für Klinische Chemie for the most outstanding analytical development in clinical chemistry worldwide during a 2-year period. In 1984 ANL awarded him its career patent leader award for the largest number of patents issued to an employee. At that time the commercial value of his inventions in terms of U.S. sales and royalties from foreign licensing were \$250 million and \$1 million, respectively. Dr. Anderson received his degrees at Duke University: a B.A. in Zoology, M.A. in Physiology, and Ph.D. in Cell Physiology. He holds 28 patents.

Constance Seniff, Vice President, Operations. Ms. Seniff has managed LSP's operations since 1993. Her background includes thirteen years in international business prior to joining LSP, five abroad in the employ of foreign firms. Ms. Seniff is responsible for helping formulate and implement business development and database commercialization strategies for LSP in coordination with the management of LSP's parent company, Large Scale Biology Corporation. Ms. Seniff has a B.Sc. degree in Business (with honors) from Florida State University..

Robert J. Walden, Vice President, Finance at LSP. Mr. Walden joined LSP in 1997 and has served as a director since 1999. He previously served as Vice President of Finance and Administration at Osiris Therapeutics, Inc., and as Chief Financial Officer at the American Type Culture Collection (ATCC). Mr. Walden received his degree in Finance from the University of Maryland.

Jean-Paul Hofmann, Ph.D., Vice President, Software Development at LSP. Dr. Hofmann is a plant geneticist by training, having earned a B.S. in Biology, M.S. in Biochemistry and Genetics, and Ph.D. in Plant Genetics from the University of Orsay, Paris. He has extensive



experience in using 2-DE in agronomic research and in designing analytical software for 1- and 2-D applications. He has held senior scientific positions in industry and research institutes, in the U.S., France and the Ivory Coast.

John Taylor, Ph.D., Vice President, Software Development and Bioinformatics. Dr. Taylor is the principal developer of Kepler™, LSP's analytical software for automated 2-DE pattern analysis. Prior to joining LSB, Dr. Taylor served as computer scientist in the Molecular Anatomy Program at Argonne, and on the research staffs of the University of Chicago and the Armed Forces Institute of Pathology in Washington, D.C. Dr. Taylor received a B.S. in Physics from the University of South Carolina, and a Ph.D. in Nuclear Physics from Duke University.

Sandra Steiner, Ph.D., currently serves as Vice President Proteomics Applications. Prior to joining the Company, Dr. Steiner founded and directed the Molecular Toxicology Group at Novartis in Basel, Switzerland and was a member in several multi-disciplinary drug development project teams. Dr. Steiner received her Ph.D. in Toxicology/Pharmacology from the University of Basel, Switzerland.

[back to index](#)

© 2000 Large Scale Biology Corporation. All Rights Reserved Worldwide.

